

# Peak Detection as Multiple Testing

Armin Schwartzman<sup>1</sup>, Yulia Gavrilov<sup>1</sup>, Robert J. Adler<sup>2</sup>

<sup>1</sup> Department of Biostatistics, Harvard School of Public Health

<sup>2</sup> Department of Electrical Eng., Technion - Israel Institute of Technology

August 12, 2010

## Abstract

This paper considers the problem of detecting equal-shaped non-overlapping unimodal peaks in the presence of Gaussian ergodic stationary noise, where the number, location and heights of the peaks are unknown. A multiple testing approach is proposed in which, after kernel smoothing, the presence of a peak is tested at each observed local maximum. The procedure provides strong control of the family wise error rate and the false discovery rate asymptotically as both the signal-to-noise ratio (SNR) and the search space get large, where the search space may grow exponentially as a function of SNR. Simulations assuming a Gaussian peak shape and a Gaussian autocorrelation function show that desired error levels are achieved for relatively low SNR and are robust to partial peak overlap. Simulations also show that detection power is maximized when the smoothing bandwidth is close to the bandwidth of the signal peaks, akin to the well-known matched filter theorem in signal processing. The procedure is illustrated in an analysis of electrical recordings of neuronal cell activity.

## 1 Introduction

Peak detection is a common statistical problem in the analysis of high-throughput data. Examples include identification of binding sites on the genome (Zhang et al., 2008), DNA sequencing (Li and Speed, 2000), identification of proteins in mass spectrometry (Morris et al., 2006; Harezlak et al., 2008), detection of action potentials in neuronal recordings (Baccus and Meister, 2002), detection of heart beats in electrocardiograms (Arzeno et al., 2008), and identification of signatures in galaxy spectra (Brutti et al., 2007). A crucial step in the analysis of these data, both for dimension reduction and consequent inference, is the detection of an unknown number of signal peaks with a temporal or spatial structure in the presence of background noise. The main challenge in these problems is that both the number of peaks and their location are unknown. In addition, the data often consist of a single long sequence with no replicates.

While there are many peak detection algorithms in the scientific literature, they tend to be geared toward specific applications and their performance is often evaluated empirically. In particular, peak detection algorithms often require a threshold, but the choice of the threshold is ad-hoc and does not take into account the error inflation produced by multiple testing. Errors in peak detection can lead to erroneous conclusions in later steps of the data analysis. There is a need to approach peak detection from a formal statistical viewpoint. Our objective is to develop a general statistical procedure for identifying signal peaks in the

presence of background noise with proven error control, while at the same time being easy to implement and efficient to run on large data sets.

In this paper, we consider the specific problem of detecting equal-shaped non-overlapping unimodal peaks in the presence of Gaussian stationary noise, where the number, locations and heights of the peaks are unknown. We have chosen this particular setting for its analytical tractability, but we believe that it captures the essence of the general peak detection problem and can serve as a theoretical basis for future applications in the particular scientific disciplines.

The assumption of not knowing the number of peaks is key. If the number of peaks were known, then the unknown locations could be estimated solving a nonlinear least-squares problem (O’Brien et al., 1994; Li and Speed, 2000, 2004). The main difficulty of this approach is that not knowing the number of peaks implies not knowing the number of location parameters to be estimated. As a consequence, the problem becomes akin to the model selection problem in regression (Li and Speed, 2000, 2004). Alternative solutions using  $L_1$  regularization include direct penalization of the estimated signal (O’Brien et al., 1994) and a modification of the LASSO algorithm where the penalty is applied to the difference between consecutive coefficients to account for the ordered structure of the data (Tibshirani et al., 2005).

Rather than an estimation problem, we view peak detection as a multiple testing problem, where, at each of a set of locations, a test is performed for whether the signal is nonzero at that location. Our approach can be viewed essentially as a search for peaks over the length of the data. The idea has been used elsewhere (Yasui et al., 2003; Morris et al., 2006; Chumbley and Friston, 2009) but not formally for multiple testing, and is motivated as follows. A full search for peaks would require testing every single observed point for significance. However, if the peaks are assumed unimodal and non-overlapping, dimensionality can be reduced dramatically by testing only at locations that resemble peaks, that is, local maxima of the observed sequence. In addition, it is known that signal-to-noise ratio (SNR) can be improved by local smoothing such as averaging over a local neighbourhood. Based on these ideas, our proposed algorithm consists of the following steps:

1. *Kernel smoothing*
2. *Candidate peaks*: find local maxima of the smoothed sequence.
3. *P-values*: compute a p-value at each local maximum, defined as the probability of peering the observed intensity of the local maximum or higher according to the distribution that would be expected if we only observed noise.
4. *Multiple testing*: use a multiple testing procedure to find a global threshold and declare significant all peaks exceeding that threshold.

For Step 4, we consider two standard multiple testing procedures: Bonferroni and Benjamini-Hochberg (BH) (Benjamini and Hochberg, 1995). Our peak detection algorithm has the advantage of being simple, easy to remember and efficient to implement for large data sets. At the same time, we believe it to be powerful and we show that it provides guaranteed global error control. As measures of global error we consider both family-wise error rate (FWER), to be controlled by the Bonferroni procedure, and false discovery rate (FDR) (Benjamini and Hochberg, 1995), to be controlled by the BH procedure. For simplicity, we concentrate on positive signals and one-sided tests, but this is not crucial.

Our proofs of error control assume that the noise is a continuous ergodic stationary Gaussian process. This assumption permits a closed form formula for computing the p-values corresponding to local maxima of the observed process. The distribution of the height of a local maximum of a Gaussian process is not Gaussian but has a heavier tail, and its computation requires careful conditioning based on the calculus of Palm probabilities (Cramér and Leadbetter, 1967; Adler et al., 2010). This is crucial to ensure that p-values are valid.

Another interesting and challenging aspect of the proof of error control is the fact that the number of tests, which is equal to the number of local maxima in a given interval, is a random quantity. Proofs of error control in the multiple testing literature usually assume that the number of tests is fixed. In our proofs, we overcome this difficulty using an asymptotic argument for large search space, so that the error behaves approximately as it would if the number of tests were equal to its expected value.

In the proofs, the asymptotics for large search space are combined with asymptotics for large SNR. The large SNR assumption helps solve the issue of identifiability of peaks, as it implies that each signal peak is represented by only one observed local maximum with probability tending to one. The asymptotic rates, however, are such that the search space is allowed to grow much faster than the SNR; exponentially faster. In this sense, we do not consider the large SNR assumption restrictive.

For concreteness, we conduct simulations assuming a Gaussian peak shape and a Gaussian autocorrelation function. Our simulations confirm that moderate values of SNR are enough to provide desired error levels. Our simulations also show that the performance is maintained under a substantial amount of overlap between neighboring peaks.

Defining detection power as the expected fraction of true peaks detected, we prove that the peak detection algorithm is consistent in the sense that its power tends to one under the above asymptotic conditions. We then use simulations to address the question of optimal bandwidth. In the above Gaussian autocorrelation model, we find that the detection power is maximized when the smoothing bandwidth is close to the bandwidth of the signal peaks, slightly larger when the noise is uncorrelated and getting smaller as the amount of autocorrelation in the noise increases. This result is similar to the well-known matched filter theorem in signal processing, which states that the SNR after smoothing is maximized when the smoothing kernel matches the shape and width of the signal (Simon, 1995; Pratt, 1991). Most noticeably, the optimal bandwidth for peak detection as multiple testing is different, in fact much larger, than the usual optimal bandwidth for nonparametric estimation. A distinction between our setting and the matched filter theorem is that in many applications of the latter in digital communications and radar screening, the time of testing is prespecified, while in our case it is random.

We illustrate our procedure with a data set of neural electrical recordings, where the objective is to detect action potentials representing cell activity (Baccus and Meister, 2002). The data analysis takes advantage of the sparsity of the signal in order to estimate the parameters of the noise process.

The rest of the paper is organized as follows. Section 2 presents the theoretical results. Section 3 presents the simulation results. Section 4 presents the data example. Section 5 summarizes. Proofs are given in Section 6. All the simulations and the data analysis were implemented in R.

## 2 Theory

### 2.1 The model

Consider the signal-plus-noise model

$$y(t) = \mu(t) + z(t), \quad t \in \mathbb{R} \quad (1)$$

where  $\mu(t)$  is the signal we wish to detect and  $z(t) \in \mathcal{C}^2$  is stationary ergodic zero-mean Gaussian noise. The signal  $\mu(t)$  is a (sparse) train of positive peaks of the form

$$\mu(t) = \sum_{j=-\infty}^{\infty} a_j h_b(t - \tau_j), \quad h_b(t) = \frac{1}{b} h\left(\frac{t}{b}\right) \quad (2)$$

where  $a_j, b > 0$ . The peak shape  $h(t) \geq 0$  is assumed unimodal with mode at  $t = 0$  and no other critical points within its support  $S = \{t : h(t) > 0\}$ , where  $S$  is assumed to be compact and connected. Assume that  $h(t)$  has unit action  $\int_{-\infty}^{\infty} h(t) dt = 1$ . Let

$$w_\gamma(t) = \frac{1}{\gamma} w\left(\frac{t}{\gamma}\right), \quad \gamma > 0$$

be a unimodal kernel with compact and connected support and  $\int_{-\infty}^{\infty} w_\gamma(t) dt = 1$ . Define

$$h_\gamma(t) = w_\gamma(t) * h_b(t) \quad (3)$$

with support  $S_\gamma = \{t : h_\gamma(t) > 0\}$ . We assume that  $h_\gamma(t)$  is unimodal with mode at some interior point of  $S$  not necessarily equal to 0, with no other critical points. In addition we assume that  $h_\gamma(t)$  is twice differentiable in the interior of  $S_\gamma$ . Note that if, for example, both  $h(t)$  and  $w(t)$  are truncated Gaussian functions, then all the assumptions made on  $h_\gamma(t)$  above are valid.

Finally, the smoothed signal and smoothed noise are defined as

$$\mu_\gamma(t) = w_\gamma(t) * \mu(t) = \sum_{j=-\infty}^{\infty} a_j h_\gamma(t - \tau_j), \quad z_\gamma(t) = w_\gamma(t) * z(t). \quad (4)$$

We require that the supports  $S_{j,\gamma} = \{t : h_\gamma(t - \tau_j) > 0\}$  do not overlap for all  $j$ . In this sense the signal can be considered sparse.

### 2.2 The procedure

Suppose we observe  $y(t)$  in the segment  $T = [-L/2, L/2]$ , which contains  $J_L$  peaks. The objective is to identify a set of locations  $\tilde{\tau}_1, \dots, \tilde{\tau}_{\tilde{m}}$  that are close in location and in number to the true set of peak locations  $\tau_1, \dots, \tau_{J_L}$ , while controlling the probability of obtaining false peaks. Consider the following procedure.

#### Procedure 2.1.

1. *Kernel smoothing:* Construct the process

$$x_\gamma(t) = w_\gamma(t) * y(t) = \int_{-\infty}^{\infty} w_\gamma(t - s) y(s) ds, \quad (5)$$

where we ignore boundary effects at  $\pm L/2$ .

2. *Candidate peaks*: Find all the local maxima of  $x_\gamma(t)$  in  $[-L/2, L/2]$ , i.e. find the set

$$\tilde{T} = \left\{ t \in \left[ -\frac{L}{2}, \frac{L}{2} \right] : \dot{x}_\gamma(t) = \frac{dx_\gamma(t)}{dt} = 0, \quad \ddot{x}_\gamma(t) = \frac{d^2x_\gamma(t)}{dt^2} < 0 \right\}. \quad (6)$$

3. *P-values*: For each  $t \in \tilde{T}$  calculate the p-value for testing the hypothesis

$$\mathcal{H}_0(t) : \mu(t) = 0 \quad \text{vs.} \quad \mathcal{H}_A(t) : \mu(t) > 0$$

4. *Multiple testing*: Apply a multiple testing procedure on the set of p-values and declare significant all peaks whose p-values are smaller than the corresponding threshold, which may depend on  $\gamma$  and  $L$ .

The idea behind Procedure 2.1 is very simple: smooth, find local maxima, feed the list of local maxima into any standard multiple testing procedure. These steps are easy to remember and implement. The first two steps are straightforward. The last two need careful study, which we do next.

Step 3 is detailed in Section 2.3 below. For Step 4, we use the Bonferroni procedure to control FWER and the BH procedure to control FDR. Let  $\tilde{m}$  be the number of local maxima in  $\tilde{T}$ , i.e. the number of tested hypotheses. To apply the Bonferroni procedure at level  $\alpha$  we compare each p-value to the threshold  $\alpha/\tilde{m}$  and reject all the hypotheses whose p-values are below the threshold. To apply the BH procedure at level  $\alpha$  we first order the p-values and then compare the ordered  $i$ -th p-value with  $i\alpha/\tilde{m}$ . Defining  $k$  as the maximal index for which the ordered p-value is smaller than the corresponding threshold, we reject  $k$  hypotheses with the  $k$  smallest p-values. More details are given in the Sections below.

### 2.3 P-values

A crucial step in implementing any marginal multiple testing procedure is calculating p-values. P-values are always computed under the complete null hypothesis of no signal anywhere. For Step 3 of Procedure 2.1, we proceed as follows.

**Definition 2.2.** Assume the model of Section 2.1 with  $\mu(t) = 0, \forall t$ , so that  $x_\gamma(t) = z_\gamma(t)$ , given by (4). Let  $F_\gamma(u)$  denote the *right* cumulative distribution function (cdf) of  $z_\gamma(t)$  at the local maxima  $t \in \tilde{T}$ ,

$$F_\gamma(u) = \mathbb{P} \left\{ z_\gamma(t) > u \mid t \in \tilde{T} \right\}. \quad (7)$$

Then the p-value of the observed  $x_\gamma(t)$  at  $t \in \tilde{T}$  is

$$p_\gamma(t) = F_\gamma[x_\gamma(t)], \quad t \in \tilde{T}. \quad (8)$$

The probability in (7) follows the Palm distribution for maxima and we use its properties to compute an explicit formula for p-values. One needs to be careful while computing this probability to avoid bias because of two reasons. First, although we use the usual symbol for conditioning, this is not the usual conditioning event. Computing this probability by usual conditioning gives the Gaussian distribution. However, computing the probabilities just at local maxima based on the Gaussian distribution will lead to the biased down p-values. Second, we are conditioning on an event of probability zero. See Adler et al. (2010, Ch. 6) for a more detailed discussion. The formula for evaluating the required distribution (7) is given by Proposition 2.3 below.

**Proposition 2.3.** *Assume the model of Section 2.1 and the procedure of Section 2.2. Under the complete null hypothesis  $\mu(t) = 0, \forall t$ , define the moments*

$$\sigma_\gamma^2 = \text{var}[z_\gamma(t)], \quad \lambda_{2,\gamma} = \text{var}[\dot{z}_\gamma(t)], \quad \lambda_{4,\gamma} = \text{var}[\ddot{z}_\gamma(t)]. \quad (9)$$

Then,

$$F_\gamma(u) = 1 - \Phi\left(u\sqrt{\frac{\lambda_{4,\gamma}}{\Delta}}\right) + \sqrt{\frac{2\pi\lambda_{2,\gamma}^2}{\lambda_{4,\gamma}\sigma_\gamma^2}}\phi\left(\frac{u}{\sigma_\gamma}\right)\Phi\left(u\sqrt{\frac{\lambda_{2,\gamma}^2}{\Delta\sigma_\gamma^2}}\right), \quad (10)$$

where

$$\Delta = \sigma_\gamma^2\lambda_{4,\gamma} - \lambda_{2,\gamma}^2 \quad (11)$$

and  $\phi(x)$ ,  $\Phi(x)$ , are the standard normal density and cdf, respectively.

This result was proven by Cramér and Leadbetter (1967, Ch. 10), using the well known Kac-Rice formula (Rice, 1945), (Adler and Taylor, 2007, Ch. 11). The proof, which we omit here, is based on other intermediate results that we will also need later, and so we state them in the next lemma.

**Lemma 2.4.** *Let  $z(t) \in \mathcal{C}^2$  be an ergodic stationary zero-mean Gaussian process with spectral moments  $\sigma^2 = \text{var}[z(t)]$ ,  $\lambda_2 = \text{var}[\dot{z}(t)]$  and  $\lambda_4 = \text{var}[\ddot{z}(t)]$ , defined on a compact set  $T \subset \mathbb{R}$  with non-empty interior and finite measure  $|T|$ .*

1. *Define respectively the number of local maxima in  $T$  and the number of local maxima in  $T$  that cross the threshold  $u$  as*

$$\begin{aligned} \tilde{m}(-\infty; T) &= \#\{t \in T : \dot{z}(t) = 0, \ddot{z}(t) < 0\} \\ \tilde{m}(u; T) &= \#\{t \in T : z(t) > u, \dot{z}(t) = 0, \ddot{z}(t) < 0\}. \end{aligned}$$

Then

$$\begin{aligned} \mathbb{E}[\tilde{m}(-\infty; T)] &= |T|\mathbb{E}[\tilde{m}(-\infty; [0, 1])], \\ \mathbb{E}[\tilde{m}(u; T)] &= |T|\mathbb{E}[\tilde{m}(u; [0, 1])], \end{aligned}$$

where  $\mathbb{E}[\tilde{m}(-\infty; [0, 1])] = \sqrt{\lambda_4/\lambda_2}/(2\pi)$  is the expected number of local maxima and  $\mathbb{E}[\tilde{m}(u; [0, 1])]$  is the expected number of local maxima above the threshold  $u$  in  $[0, 1]$ , with expression given in Cramér and Leadbetter (1967, Ch. 10).

2. *The right cdf of the heights of the local maxima of  $z(t)$  is given by*

$$\mathbb{P}\left\{z(t) > u \mid \dot{z}(t) = 0, \ddot{z}(t) < 0\right\} = \frac{\mathbb{E}[\tilde{m}(u; [0, 1])]}{\mathbb{E}[\tilde{m}(-\infty; [0, 1])]}. \quad (12)$$

Proposition 2.3 follows directly from Lemma 2.4 applied to the process  $z_\gamma(t)$ . In particular, (10) follows directly from evaluating (12). Referring to (10), note that, for large enough  $u$ ,

$$F_\gamma(u) \approx \sqrt{\frac{2\pi\lambda_{2,\gamma}^2}{\lambda_{4,\gamma}\sigma_\gamma^2}}\phi\left(\frac{u}{\sigma_\gamma}\right). \quad (13)$$

Therefore, the tail of the distribution is proportional to a Gaussian density, so it behaves like a Rayleigh distribution.

A peculiar characteristic of Procedure 2.1 is that the number of tests  $\tilde{m}$ , which is the same as the number of p-values, is random. In particular, under the complete null hypothesis

$\mu(t) = 0, \forall t$ , the expected number of tests can be computed explicitly applying Lemma 2.4 to the process  $z_\gamma(t)$  as

$$\mathbb{E}[\tilde{m}(-\infty; [-L/2, L/2])] = \frac{L}{2\pi} \sqrt{\frac{\lambda_{4,\gamma}}{\lambda_{2,\gamma}}}. \quad (14)$$

The quantities  $\sigma_\gamma^2$ ,  $\lambda_{2,\gamma}$ , and  $\lambda_{4,\gamma}$  in Proposition 2.3 depend on the kernel  $w_\gamma(t)$  and the autocorrelation of the original noise process  $z(t)$ . A specific form of the p-values and the expected number of tests may be obtained for the following Gaussian autocorrelation model.

**Example 2.5 (Gaussian autocorrelation model).** Assume the complete null hypothesis. Suppose  $w(t) = \phi(t)\mathbf{1}[-c, c]$  is a truncated normal density and assume that the autocovariance function of the noise process  $z(t)$  is proportional to a normal density, i.e.

$$z(t) = \sigma \int_{-\infty}^{\infty} w_\nu(s - t) dB(s), \quad w_\nu(t) = \frac{1}{\nu} \phi\left(\frac{t}{\nu}\right)$$

where  $B(s)$  is standard Brownian motion and  $\nu > 0$ . Ignoring the truncation at  $\pm c$ , the required moments in (10) or (14) are

$$\sigma_\gamma^2 = \frac{\sigma^2}{2\sqrt{\pi}\xi}, \quad \lambda_{2,\gamma} = \frac{\sigma^2}{4\sqrt{\pi}\xi^3}, \quad \lambda_{4,\gamma} = \frac{3\sigma^2}{8\sqrt{\pi}\xi^5}, \quad \xi = \sqrt{\gamma^2 + \nu^2}. \quad (15)$$

Derivations are given in Section 6.1. Therefore, for the Gaussian autocorrelation model,

$$F_\gamma(u) = 1 - \Phi\left(\frac{u}{\sigma} \sqrt{3\sqrt{\pi}\xi}\right) + \sqrt{\frac{2\pi}{3}} \phi\left(\frac{u}{\sigma} \sqrt{2\sqrt{\pi}\xi}\right) \Phi\left(\frac{u}{\sigma} \sqrt{\sqrt{\pi}\xi}\right), \quad (16)$$

$$\mathbb{E}[\tilde{m}(-\infty; [-L/2, L/2])] = \frac{L}{2\pi\xi} \sqrt{\frac{3}{2}}.$$

If  $w(t)$  is truncated at  $\pm c$  where  $c$  is large enough, say  $c = 4$ , the moments in (15) are a good approximation for the same moments computed in the truncated version.

## 2.4 Error rates definitions

In this subsection we define the two error rates to control, FWER and FDR. Let us define first what is considered a false discovery and what is considered a true discovery under the true model.

**Definition 2.6.** Let  $S_j = \{t : h_b(t - \tau_j) > 0\}$  be the support of  $h_b(t - \tau_j)$ . Define the signal region  $\mathbb{S}_1$  and null region  $\mathbb{S}_0$  respectively by

$$\mathbb{S}_1 = \bigcup_{j=1}^J S_j \quad \text{and} \quad \mathbb{S}_0 = [-L/2, L/2] \setminus \left( \bigcup_{j=1}^J S_j \right).$$

It is known that kernel smoothing expands the signal support. This effect requires care as it increases the probability of obtaining false positives in the regions neighboring the signal (Pacífico et al., 2004b). The larger the bandwidth  $\gamma$ , the greater the distortion. In particular, the support defined with respect to  $w_\gamma$  is larger than  $S_j$ , and we define it formally next.

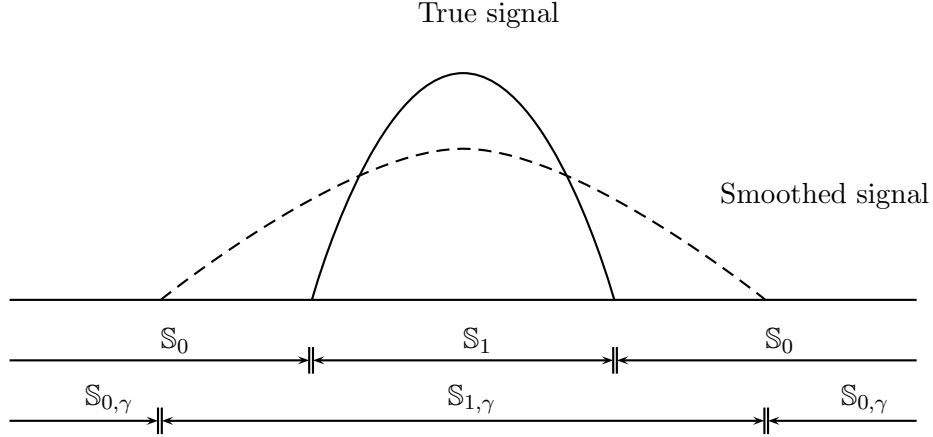


Figure 1: True and smoothed signal.

**Definition 2.7.** Fix  $\gamma > 0$  and define  $S_{j,\gamma} = \{t : h_\gamma(t - \tau_j) > 0\}$  where  $h_\gamma(t)$  is given by (3). Define the expanded signal region  $\mathbb{S}_{1,\gamma}$  and reduced null region  $\mathbb{S}_{0,\gamma}$  respectively by

$$\mathbb{S}_{1,\gamma} = \bigcup_{j=1}^J S_{j,\gamma} \quad \text{and} \quad \mathbb{S}_{0,\gamma} = [-L/2, L/2] \setminus \left( \bigcup_{j=1}^J S_{j,\gamma} \right).$$

The above definitions are illustrated schematically in Figure 1. Some useful equalities are: (1)  $\mathbb{S}_0 \cup \mathbb{S}_1 = \mathbb{S}_{0,\gamma} \cup \mathbb{S}_{1,\gamma}$  and (2)  $\mathbb{S}_{1,\gamma} \setminus \mathbb{S}_1 = \mathbb{S}_0 \setminus \mathbb{S}_{0,\gamma}$ . We refer to the latter set, the difference between the expanded signal support and the true signal support, as the transition region.

We follow the classical view of error definitions. A significant local maximum that belongs to  $\mathbb{S}_0$  is considered an error. A significant local maximum that belongs to  $\mathbb{S}_1$  is considered correct. Moreover, if more than one local maximum occur within the same peak, all significant local maxima are considered as correct. However, all these local maxima are counted as one peak for the purposes of power, defined formally below in Section 2.8. Thus power is not inflated. Nevertheless, this situation does not affect the theoretical results because under our asymptotic assumptions (Theorem 2.12) each peak is represented by one local maximum with probability tending to 1.

To simplify the notation, we define the number of falsely rejected local maxima using the threshold  $u$  as:

$$V(u) = \#\{t \in \tilde{T} \cap \mathbb{S}_0 : x_\gamma(t) > u\} \quad \text{and} \quad V_\gamma(u) = \#\{t \in \tilde{T} \cap \mathbb{S}_{0,\gamma} : x_\gamma(t) > u\},$$

where  $V(u)$  counts false discoveries in the full null region and  $V_\gamma(u)$  counts them only in the null region minus the transition region. Similarly, we define the number of correctly rejected hypotheses using the threshold  $u$  as

$$W(u) = \#\{t \in \tilde{T} \cap \mathbb{S}_1 : x_\gamma(t) > u\} \quad \text{and} \quad W_\gamma(u) = \#\{t \in \tilde{T} \cap \mathbb{S}_{1,\gamma} : x_\gamma(t) > u\},$$

where  $W(u)$  counts discoveries in the true signal region and  $W_\gamma(u)$  counts them in the signal region plus the transition region. The total number of rejections using threshold  $u$  is

$$R(u) = V(u) + W(u) = \#\{t \in \tilde{T} : x_\gamma(t) > u\}.$$



The number of tests where the null hypothesis is true is denoted by

$$\tilde{m}_0 = V(-\infty) = \#\{t \in \tilde{T} \cap \mathbb{S}_0\},$$

and the total number of tests is

$$\tilde{m} = R(-\infty) = \#\{t \in \tilde{T}\}.$$

The number of tests where the null hypothesis is false is denoted by  $\tilde{m}_1 = \tilde{m} - \tilde{m}_0$ .

**Definition 2.8.** Define the family-wise error rate (FWER) for any threshold  $u$  as the probability that there exists at least one local maximum of  $x_\gamma(t)$  in the null region above the threshold  $u$ :

$$\text{FWER}(u) = \mathbb{P}\{V(u) \geq 1\} = \mathbb{P}\left\{\tilde{T} \cap \mathbb{S}_0 \neq \emptyset \text{ and } \max_{t \in \tilde{T} \cap \mathbb{S}_0} x_\gamma(t) > u\right\}. \quad (17)$$

**Definition 2.9.** Define the false discovery rate (FDR) as the expected proportion of falsely rejected hypotheses using a fixed threshold  $u$ :

$$\text{FDR}(u) = \mathbb{E}\left\{\frac{V(u)}{R(u) \vee 1}\right\}. \quad (18)$$

Note that if  $\tilde{T}$  is empty then  $\tilde{m} = 0$  and  $V(u) = 0$  and so  $\text{FWER}(u) = \text{FDR}(u) = 0$ . However, the probability of this event goes to zero as  $L$  increases.

## 2.5 Weak control of FWER

In this subsection we assume the complete null hypothesis  $\mu(t) = 0, \forall t \in [-L/2, L/2]$ . Note that under this assumption  $\tilde{m} = \tilde{m}_0$ .

In Procedure 2.1, after Step 2 one has a list of  $\tilde{m}$  local maxima. The Bonferroni procedure at level  $\alpha$  defines a threshold of the form  $\alpha/\tilde{m}$  and rejects all hypotheses whose p-values are below this threshold. Note that in the usual Bonferroni procedure the number of tested hypotheses is constant. In our case the number of tested hypotheses is random, therefore we consider two versions. We refer to the version in part (2) of Theorem 2.10 as the Bonferroni procedure, while the version in part (1) can be viewed as the 'limit' of the Bonferroni procedure.

**Theorem 2.10.** Assume the model of Section 2.1 and the procedure of Section 2.2. Let  $\mu(t) = 0, \forall t$ . Fix  $\alpha > 0$ .

1. Suppose the null hypothesis  $\mathcal{H}_0(t)$  is rejected at  $t \in \tilde{T}$  if

$$p_\gamma(t) < \frac{\alpha}{\mathbb{E}[\tilde{m}]} \iff x_\gamma(t) > u_{\text{Bon}}^* = F_\gamma^{-1}\left(\frac{\alpha}{\mathbb{E}[\tilde{m}]}\right). \quad (19)$$

Then  $\text{FWER}(u_{\text{Bon}}^*) \leq \alpha$ .

2. Suppose the null hypothesis  $\mathcal{H}_0(t)$  is rejected at  $t \in \tilde{T}$  by a Bonferroni procedure with  $\tilde{m}$  tests, i.e. if

$$p_\gamma(t) < \frac{\alpha}{\tilde{m}} \iff x_\gamma(t) > \tilde{u}_{\text{Bon}} = F_\gamma^{-1}\left(\frac{\alpha}{\tilde{m}}\right). \quad (20)$$

If  $\tilde{m} = 0$  we define  $\alpha/\tilde{m}$  as infinity. Then  $\limsup_{L \rightarrow \infty} \text{FWER}(\tilde{u}_{\text{Bon}}) \leq \alpha$ .

The proof of Theorem 2.10 is given in Section 6.2. The first part of the theorem is not asymptotic and its proof is a direct consequence of the definition of p-values. The threshold  $u_{\text{Bon}}^*$  in (19) is deterministic. For example, in the Gaussian autocorrelation model this threshold can be computed by substituting (16). The threshold  $\tilde{u}_{\text{Bon}}$  in (20) depends on the random quantity  $\tilde{m}$  and is equivalent to applying the Bonferroni procedure on the random set of local maxima,  $\tilde{T}$ . The proof of the second part is based on the fact that by the weak law of large numbers  $\tilde{m}$  is close to its expectation  $\mathbb{E}[\tilde{m}]$  for large enough  $L$ .

Using (13), the thresholds in Theorem 2.10 may be approximated by

$$\begin{aligned} u_{\text{Bon}}^* &\approx \sigma_\gamma \phi^{-1} \left( \frac{\alpha}{L} \sqrt{\frac{2\pi\sigma_\gamma^2}{\lambda_{2,\gamma}}} \right) \approx \sigma_\gamma \sqrt{2 \log \left( \frac{L}{\alpha} \sqrt{\frac{\lambda_{2,\gamma}}{2\pi\sigma_\gamma^2}} \right)} \\ \tilde{u}_{\text{Bon}} &\approx \sigma_\gamma \phi^{-1} \left( \frac{\alpha}{\tilde{m}_L} \sqrt{\frac{\lambda_{4,\gamma}\sigma_\gamma^2}{2\pi\lambda_{2,\gamma}^2}} \right) \approx \sigma_\gamma \sqrt{2 \log \left( \frac{\tilde{m}_L}{\alpha} \sqrt{\frac{2\pi\lambda_{2,\gamma}^2}{\lambda_{4,\gamma}\sigma_\gamma^2}} \right)} \end{aligned}$$

Notice that asymptotically both thresholds have a similar form to the universal threshold of Donoho and Johnstone (1994, 1995).

## 2.6 Strong control of FWER

Let us start with brief reminder about the assumptions of the true model. We observe  $y(t)$  as defined in (1) in the segment  $T = [-L/2, L/2]$ , where the signal contains  $J_L$  peaks at locations  $\tau_j \in S_j$  for  $1 \leq j \leq J_L$ , where  $S_j$  is the finite support of the  $j$ th peak. Recall that  $h_\gamma(t) = w_\gamma(t) * h(t)$  is the peak shape after smoothing. It achieves its supremum at a single point  $\tau_{j,\gamma} \in S_j$  which is not necessarily the same as  $\tau_j$  and has no other critical points. We define the signal after smoothing as

$$\mu_\gamma(t) = \sum_{j=1}^{J_L} a_j h_\gamma(t - \tau_{j,\gamma}). \quad (21)$$

The next theorem describes the strong control of FWER for this model.

**Theorem 2.11.** *Assume the model of Section 2.1 and the procedure of Section 2.2. Fix  $\alpha > 0$ .*

1. *Suppose the null hypothesis  $\mathcal{H}_0(t)$  is rejected at  $t \in \tilde{T}$  if*

$$p_\gamma(t) < \frac{\alpha}{\mathbb{E}[\tilde{m}]} \iff x_\gamma(t) > u_{\text{Bon}}^* = F_\gamma^{-1} \left( \frac{\alpha}{\mathbb{E}[\tilde{m}]} \right). \quad (22)$$

*Then, for all  $L$ ,  $\limsup \text{FWER}(u_{\text{Bon}}^*) \leq \alpha$ , as  $a_j \rightarrow \infty, \forall j$ .*

2. *Suppose the null hypothesis  $\mathcal{H}_0(t)$  is rejected at  $t \in \tilde{T}$  by the Bonferroni procedure with  $\tilde{m}$  tests, i.e. if*

$$p_\gamma(t) < \frac{\alpha}{\tilde{m}} \iff x_\gamma(t) > \tilde{u}_{\text{Bon}} = F_\gamma^{-1} \left( \frac{\alpha}{\tilde{m}} \right). \quad (23)$$

*If  $\tilde{m} = 0$  we define  $\alpha/\tilde{m}$  as infinity. Then  $\limsup \text{FWER}(\tilde{u}_{\text{Bon}}) \leq \alpha$ , as  $L \rightarrow \infty$  and  $a_j \rightarrow \infty, \forall j$ , such that for all  $j$ ,  $La_j\phi(Ka_j) \rightarrow 0$  for any constant  $K$ .*

Notice that the result in the first part of Theorem 2.11 is asymptotic, in contrast to the result in the first part of Theorem 2.10. In addition,  $u_{\text{Bon}}^*$  defined in (22) cannot be computed, even assuming the Gaussian autocorrelation model, since the true signal region is unknown. In practice we recommend to apply the Bonferroni procedure defined in the second part of Theorem 2.11, which can always be implemented and provides asymptotic strong control of FWER. The rates in Theorem 2.11 part (2) may be achieved, for instance, if  $L$  increases exponentially with  $a_j$ .

The proof of Theorem 2.11 is given in Section 6.3. There we show first that the expected number of local maxima in the transition region  $\mathbb{S}_{1,\gamma} \setminus \mathbb{S}_1$  converges to 0 (Lemma 6.3). Then we use arguments similar to those in the proof of Theorem 2.10.

## 2.7 Control of FDR

In many applied problems the expected ratio of false discoveries is a more appropriate error rate to control. The next theorem states that applying the BH procedure on the random set of local maxima controls the FDR asymptotically.

**Theorem 2.12.** *Assume the model of Section 2.1 and the procedure of Section 2.2. Let  $x_\gamma^{(i)}(t)$  be the  $i$ th ordered value of  $\tilde{T}$  in ascending order. Let*

$$\tilde{u}_{\text{BH}} = u_k, \quad k = \min_{1 \leq i \leq \tilde{m}} \{i : x_\gamma^{(i)}(t) > u_i\} \quad (24)$$

*be the threshold obtained by applying the BH procedure at the level  $\alpha$  on the random set  $\tilde{T}$  with  $\tilde{m} \neq 0$  tests, where*

$$u_i = F_\gamma^{-1} \left( \frac{\tilde{m} - i + 1}{\tilde{m}} \alpha \right), \quad 1 \leq i \leq \tilde{m}$$

*are the constants corresponding to the BH procedure. Reject  $\tilde{m} - k + 1$  hypotheses corresponding to  $x_\gamma(t) > \tilde{u}_{\text{BH}}$ . If such  $k$  does not exist or  $\tilde{m} = 0$  reject nothing. Assume that for  $L \rightarrow \infty$  and  $a_j \rightarrow \infty$  for all  $j$ ,*

1. *the number of peaks  $J_L$  increases at the same rate as  $L$  or slower, i.e.  $J_L/L \rightarrow A_1$ , where  $0 < A_1 < 1$ .*
2.  *$L \rightarrow \infty$  and  $a_j \rightarrow \infty, \forall j$  such that  $\forall j, L\phi(Ka_j^\delta) \rightarrow 0$  for all  $\delta > 0$  and any constant  $K$ .*

*Then,*

$$\limsup \text{FDR}(\tilde{u}_{\text{BH}}) \leq \alpha.$$

The rates of this theorem may be achieved if  $L$  grows exponentially with  $a_j$ . The proof is given in Section 6.4. First we show that: (1) a local maximum exists in a small neighbourhood of a true peak mode with probability tending to 1; (2) this local maximum is rejected for any fixed threshold with probability tending to 1 (Lemma 6.4). The proof of the theorem is then based on the following arguments. It is known that the threshold of the BH procedure can be viewed as the largest solution of the equation  $\alpha G(u) = F_\gamma(u)$ , where  $G(u)$  is the empirical right cumulative distribution function of  $x_\gamma(t)$ ,  $t \in \tilde{T}$  (Genovese et al., 2002). The ergodic assumption guarantees that  $G(u)$  has a limit. Replacing  $G(u)$  by its limit we find that the asymptotic solution  $u_{\text{BH}}^*$  satisfies

$$F_\gamma(u_{\text{BH}}^*) = \frac{\alpha A_1}{A_1 + \mathbb{E}[\tilde{m}_{0,\gamma}; [0, 1]](1 - \alpha)}. \quad (25)$$

We show that under the conditions of the theorem, the threshold  $u_{\text{BH}}^*$  asymptotically controls the FDR below the desired level  $\alpha$ . Since  $F_\gamma(\tilde{u}_{\text{BH}}) \rightarrow F_\gamma(u_{\text{BH}}^*)$ , the FDR level will be asymptotically controlled as well when using  $\tilde{u}_{\text{BH}}$  instead of  $u_{\text{BH}}^*$ .

Notice that, in contrast to the Bonferroni procedure, where the deterministic threshold  $u_{\text{Bon}}^*$  in (22) grows unbounded with increasing  $L$ , the asymptotic threshold for the BH procedure  $u_{\text{BH}}^*$  in (25) is finite, and depends on the asymptotically fixed proportion of false null hypotheses,  $A_1$ .

## 2.8 Power

We define the statistical power of Procedure 2.1 as the expected fraction of true discovered peaks:

$$\begin{aligned} \text{Power}(u) &= \mathbb{E} \left[ \frac{1}{J_L} \sum_{j=1}^{J_L} 1 \left( \tilde{T} \cap S_j \neq \emptyset \text{ and } \max_{\tilde{t} \in \tilde{T} \cap S_j} x_\gamma(\tilde{t}) > u \right) \right] \\ &= \mathbb{P} \left\{ \tilde{T} \cap S_j \neq \emptyset \text{ and } \max_{t \in \tilde{T} \cap S_j} x_\gamma(t) > u \right\} \end{aligned} \quad (26)$$

We discuss power in two ways. First, we show that the Bonferroni and BH procedures in (23) and (24) respectively are consistent in the sense that the power tends to 1 for any fixed value of the smoothing parameter  $\gamma$ . Second, we discuss what value of the smoothing parameter  $\gamma$  maximizes the power for finite samples.

### 2.8.1 Power of the Bonferroni and BH procedures

The consistency of both procedures is given in the next theorem.

**Theorem 2.13.**

1. *Under the conditions of Theorem 2.11 the power of the Bonferroni procedure (23) converges in probability to 1.*
2. *Under the conditions of Theorem 2.12 the power of the BH procedure (24) converges in probability to 1.*

The proof of Theorem 2.13 is based on the next lemma.

**Lemma 2.14.**

1. *Under the conditions of Theorem 2.11,  $u_{\text{Bon}}^*/[a_j h_\gamma(0)] \rightarrow 0$  in probability.*
2. *Under the conditions of Theorem 2.12,  $u_{\text{BH}}^*/[a_j h_\gamma(0)] \rightarrow 0$  in probability.*

As  $L$  and  $a_j$ ,  $j = 1, \dots, m$ , go to infinity, the deterministic Bonferroni threshold (22) goes to infinity as well because  $\mathbb{E}[\tilde{m}]$  goes to infinity. Lemma 2.14 states that  $a_j h_\gamma(0)$  goes to infinity faster than  $u_{\text{Bon}}^*$  does. In contrast, the asymptotic BH threshold (25) does not depend on  $L$  but only on the asymptotically fixed proportion of false null hypotheses,  $A_1$ . Therefore it does not increase with  $L$  but is asymptotically constant.

Note that the statements in Lemma 2.14 are about deterministic thresholds, while the statements of Theorem 2.13 are about random thresholds. As in the proofs of the previous

theorems, we use the fact that the gap between the deterministic and random thresholds for both the Bonferroni and BH procedures goes to 0.

It is known that in general, if there exists a signal anywhere, the power of the BH procedure is larger than the power of Bonferroni procedure (Benjamini and Hochberg, 1995). This is also true in our case with respect to  $u_{\text{Bon}}^*$  and  $u_{\text{BH}}^*$ . To see this, note that for any fixed and large enough  $L$  the thresholds can be approximated by

$$F_\gamma(u_{\text{Bon}}^*) = \frac{\alpha}{\mathbb{E}[\tilde{m}]} \approx \frac{\alpha}{\tilde{m}_{1,\gamma} + \mathbb{E}[\tilde{m}_{0,\gamma}]} \quad (27)$$

and

$$F_\gamma(u_{\text{BH}}^*) = \frac{\alpha A_1}{A_1 + \mathbb{E}[\tilde{m}_{0,\gamma}; [0, 1]](1 - \alpha)} \approx \frac{\alpha \tilde{m}_{1,\gamma}}{\tilde{m}_{1,\gamma} + (1 - \alpha)\mathbb{E}[\tilde{m}_{0,\gamma}]} \quad (28)$$

It immediately follows that if  $\tilde{m}_{1,\gamma} \geq 1$ , the threshold  $u_{\text{Bon}}^*$  is larger than the threshold  $u_{\text{BH}}^*$ , promising a larger power for the BH procedure.

### 2.8.2 Optimal choice of $\gamma$

Here we discuss the best choice of  $\gamma$ , i.e. the value of  $\gamma$  that maximizes the power (26). To maximize the probability of local maxima to exceed a given threshold  $u$  under the true model is a difficult problem. Therefore, we turn to less formal discussion in this section.

Lemma 6.4 in Section 6.4 shows that for any single true peak with mode at  $t = 0$ , a local maximum exists in a small neighbourhood  $I_{j,\varepsilon}$  of  $t = 0$  with probability tending to 1. The power may be approximated as

$$\begin{aligned} \text{Power}(u) &= \mathbb{P} \left\{ \tilde{T} \cap S_j \neq \emptyset \text{ and } \max_{t \in \tilde{T} \cap S_j} x_\gamma(t) > u \right\} \\ &= \mathbb{P} \left\{ \tilde{T} \cap I_{j,\varepsilon} \neq \emptyset \text{ and } \max_{t \in \tilde{T} \cap I_{j,\varepsilon}} x_\gamma(t) > u \right\} \\ &\quad + \mathbb{P} \left\{ \tilde{T} \cap (S_j \setminus I_{j,\varepsilon}) \neq \emptyset \text{ and } \max_{t \in \tilde{T} \cap (S_j \setminus I_{j,\varepsilon})} x_\gamma(t) > u \right\} \\ &\approx \mathbb{P} \left\{ \tilde{T} \cap I_{j,\varepsilon} \neq \emptyset \text{ and } \max_{t \in \tilde{T} \cap I_{j,\varepsilon}} x_\gamma(t) > u \right\} \\ &\approx \mathbb{P} \left\{ \max_{t \in I_{j,\varepsilon}} x_\gamma(t) > u \mid \text{a local maximum exists in } I_{j,\varepsilon} \right\}. \end{aligned}$$

Therefore, heuristically, the optimal value of  $\gamma$  should be close to the value of  $\gamma$  that maximizes  $\mathbb{P}(x_\gamma(0) > u)$ . This value of  $\gamma$  is approximately given by:

$$\begin{aligned} \arg\max_{\gamma} \mathbb{P}(x_\gamma(0) > u) &= \arg\max_{\gamma} \Phi \left( \frac{a_j h_\gamma(0) - u}{\sigma_\gamma} \right) = \arg\max_{\gamma} \left( \frac{h_\gamma(0)}{\sigma_\gamma} - \frac{u}{a_j \sigma_\gamma} \right) \\ &\approx \arg\max_{\gamma} \frac{h_\gamma(0)}{\sigma_\gamma} = \arg\max_{\gamma} \frac{\int_{-\infty}^{\infty} w_\gamma(s) h_b(s) ds}{\sigma \sqrt{\int_{-\infty}^{\infty} w_\gamma^2(s) ds}}, \end{aligned} \quad (29)$$

where  $\sigma$  and  $\sigma_\gamma$  are the standard deviations of the observed and smoothed processes  $y(t)$  and  $x_\gamma(t)$ , respectively. The approximation may be justified for both procedures by Lemma 2.14. In the last expression in (29), the optimal value of  $\gamma$  is that which makes  $w_\gamma(t)$  closest

to  $h_b(t)$  in an  $L^2$  sense. In particular, if  $w(t) = h(t)$ , then the optimal value of  $\gamma$  is  $b$ . This result is similar to the well-known matched filter theorem for detecting a single signal peak of known shape at a fixed time  $t$ .

**Example 2.15 (Gaussian autocorrelation model).** Suppose  $h(t) = w(t) = \phi(t)$ , the standard normal density, and assume that the autocovariance function of the noise process  $z(t)$  is proportional to a normal density as in Example 2.5. In this case

$$\operatorname{argmax}_{\gamma} \frac{h_{\gamma}(0)}{\sigma_{\gamma}} = \begin{cases} \sqrt{b^2 - 2\nu^2}, & \nu < b/\sqrt{2} \\ 0, & \nu > b/\sqrt{2} \end{cases} \quad (30)$$

We show in the simulations below that the optimal  $\gamma$  is indeed close to the one obtained by eq. (30).

## 3 Simulation Studies

### 3.1 Performance for finite space and SNR

A simulation study was carried out to investigate the FDR and FWER levels as well as the power of the proposed algorithm based on the Bonferroni and the BH procedures for a finite range  $L$  and for a finite SNR. Our simulation study is based on the model described in Section 2.1 and Example 2.5, with 20 equally spaced and non overlapping peaks of the same height. The peak shape,  $h(t)$ , is the normal density with standard deviation  $b = 3$ , truncated at  $c = \pm 2b$ . Peaks were centered at  $\tau_j = 100j - 50$ , for  $j = 1, \dots, 20$  and  $a_j$  in (2) for all  $j$  is chosen to be 10 and 15 in two different scenarios, giving a moderate and a strong SNR respectively. The noise is a stationary zero-mean Gaussian process as defined in Example 2.5 with  $\sigma = 1$  and  $\nu = 0, 1$  and 2. We sampled the time axis at  $L = 2000$  equally spaced points. Figure 2 presents a fragment of the simulated data. Note that the height of the peaks is  $a_j/\sqrt{2\pi b^2}$ .

Following Procedure 2.2, we constructed the process  $x_{\gamma}(t) = y(t) * w_{\gamma}(t)$ , where  $w_{\gamma}(t)$  is a Gaussian kernel and  $\gamma$  was equally spaced between 1 to 6.5 in steps of 0.5. For each configuration we computed the power and the error level based on 10,000 replications. The FWER level was computed as the proportion of replications for which at least one false discovery occurred. The FDR level was computed as the mean proportion of false discoveries out of the total number of discoveries. The power is presented as the mean number of truly discovered peaks out of the 20 peaks.

Figure 3 presents the FWER and FDR levels of the Bonferroni and the BH procedures for all the studied configurations. The nominal level for both procedures was set to 0.05. For relatively large values of  $\gamma$  the Bonferroni procedure may exceed the prespecified error level. This happens due to the broadening of the signal: many of the rejected local maxima correspond to real peaks, but due to large value of  $\gamma$ , the modes are shifted away from the true mode locations to the transition region, and the discoveries are no longer considered as correct. FDR is less affected because FDR is an expected ratio, so larger  $\gamma$  also results in fewer discoveries in the transition region. The effect of the broadening of the signal is less for both procedures for larger SNR. Note that for any finite SNR one can choose large enough  $\gamma$ , so the phenomenon of the broadening of the signal will lead to exceedance of the error rate. Recall, however, that asymptotically there should be no local maxima in the transition region. There is no strong dependency on SNR and on the noise autocorrelation parameter  $\nu$ .

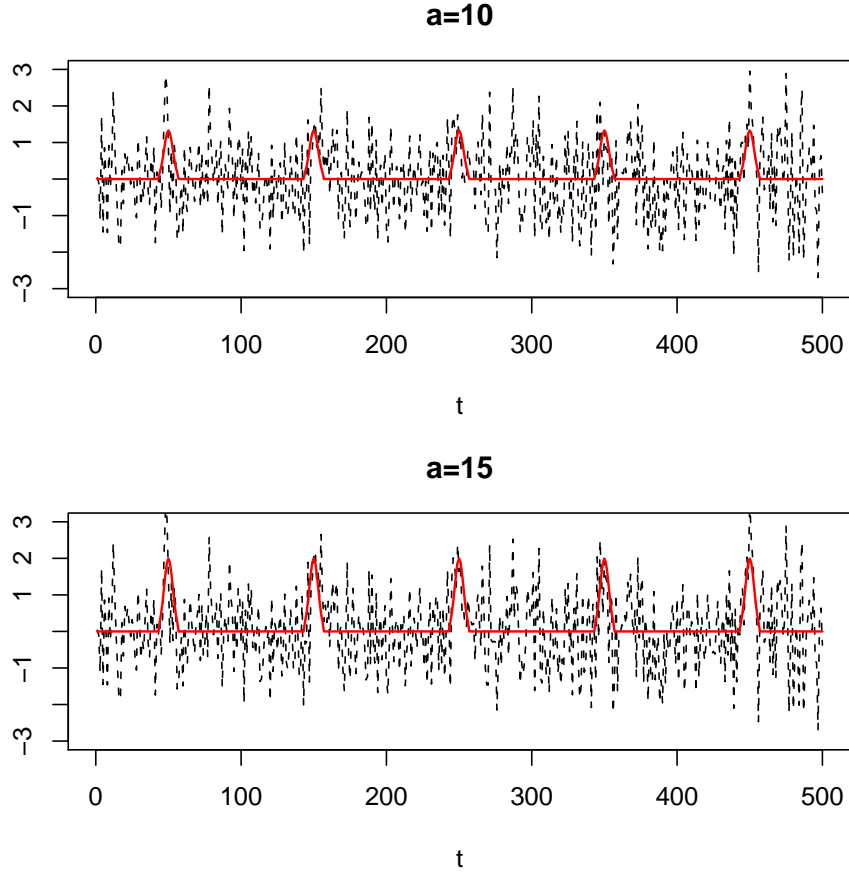


Figure 2: A fragment of the simulated data.

Figure 4 presents the power of the Bonferroni and the BH procedures, respectively, for all studied configurations. As expected, (1) the power of the BH procedure is greater than that of the Bonferroni procedure for all studied configurations; (2) the power of both procedures is greater for larger SNR. For large SNR the power of the BH procedure is almost a flat function of  $\gamma$ . This means that for large SNR the selection of  $\gamma$  is not very important as long as it is near the signal width  $b$ .

Figure 4 shows that there is one value of  $\gamma$  in each configuration for which the power is maximized. The optimal value is usually around  $b = 3$ , but it depends on the parameter  $\nu$ . The larger  $\nu$ , the smaller the value of the optimal  $\gamma$  as expected from (30). Note that the error level is always controlled if one selects the  $\gamma$  value close to the optimal one. To get more precise results about the optimal value of  $\gamma$  we performed an additional simulation study where  $\gamma$  takes values in the range 1 to 3.5 in steps of 0.1. The empirical optimal values of  $\gamma$ , as well as the optimal  $\gamma$  from equation (30), are summarized in Table 1.

As explained in Section 2.7, each peak is represented by no more than one local maximum with probability tending to 1 as the SNR increases. In our simulation we computed the probability of obtaining more than one local maximum within the same peak in the finite setting. Figure 5 shows that this probability is a decreasing function of  $\gamma$ , and a decreasing function of the SNR. Note that near the optimal value of  $\gamma$  this probability is negligible.

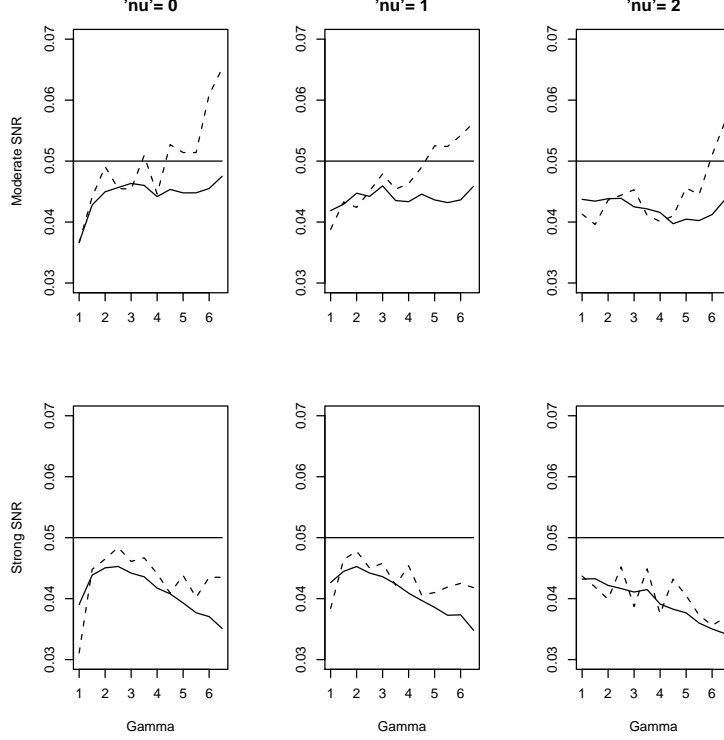


Figure 3: FDR level of the BH procedure (solid), FWER level of the Bonferroni procedure (dashed).

### 3.2 Overlapping peaks

In the second part of our simulation we test the performance of our algorithm for a non-sparse signal. In the previous simulation, the distance between two adjacent peak locations  $\tau_j$ , which we denote by  $D$ , was set to 100. We now test how the distance between the peaks  $D$  affects the power and the error level of both procedures.

For the case where  $a_j = 10$ ,  $\nu = 0$  and  $\gamma = 3.2$  (which is optimal for the Bonferroni procedure) and  $\gamma = 2.0$ , we reduce  $D$  and  $L$  while keeping fixed the proportion between the signal region,  $\mathbb{S}_1$  and the total tested region,  $\mathbb{S}$ . The first row in Figure 6 presents the error levels of the Bonferroni and BH procedures for different values of  $D$ . Recall that the support of each peak has a length of 13 ( $\tau \pm 2b$ ), so peaks start to overlap for  $D < 4b = 12$ . It can be seen that the error levels of both procedures are controlled regardless of the value of  $D$ .

Overlapping of peaks requires care of the definition of power. In this section the power was defined as the number of truly rejected peaks, where ‘truly’ means that the rejected local maximum belongs to the peak support. Since any rejected local maximum must represent just one peak, we divide the overlapping part of the support to two equal parts and add the left half to the rejection region of the peak from the left and the right half to the rejection region of the peak from the right. Thus, for any single peak the rejection region consists of the non-overlapping part of the support (in the middle of the peak) and half of each one of the overlapping parts. This makes the rejection region smaller than the original support.

The power of both procedures almost does not change when  $D$  is reduced until there



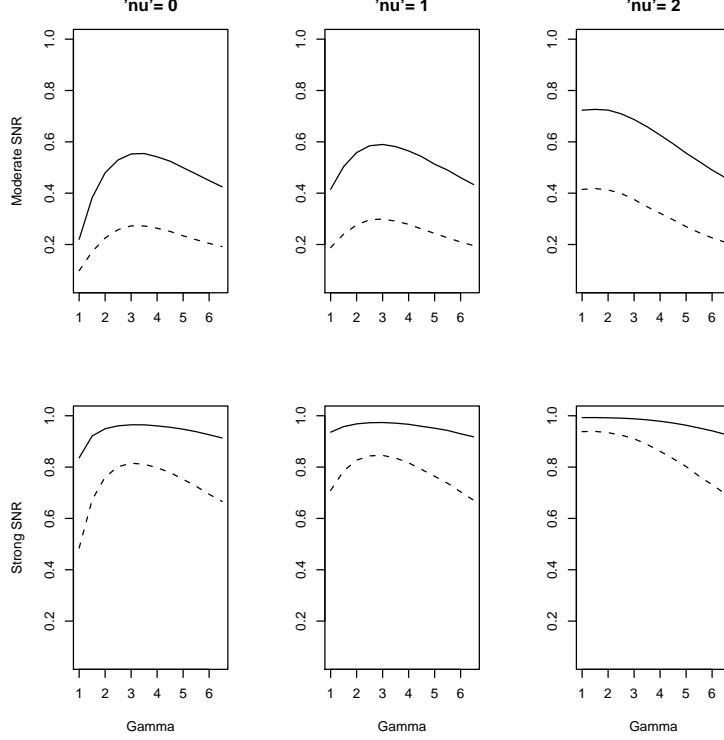


Figure 4: Power of the BH (solid) and Bonferroni (dashed) procedures.

	$a_j = 10$			$a_j = 15$		
$\nu =$	0	1	2	0	1	2
equation (30)	3.0	2.8	1.0	3.0	2.8	1.0
Bon	3.3	2.8	1.3	2.9	2.9	1.3
BH	3.2	3.1	1.3	3.4	3.0	1.2

Table 1: The ‘optimal’ value of  $\gamma$ .

is overlap in the supports; see the second row in Figure 6. Because peaks in overlapping positions are combined, the number of local maxima in the  $\mathbb{S}_1$  region decreases. However, these local maxima become more significant and are usually rejected by both the Bonferroni and the BH procedures. In the configuration presented here for  $\gamma = 3.2$ , for moderate overlapping (around 50%) 20 peaks are represented on average by 8 local maxima, while for non-overlapping supports in the same configuration 20 peaks are presented by 19.5 local maxima. In the latter case the mean number of rejections by the Bonferroni procedure is around 5 and the mean number of rejections by the BH procedure is around 11. This explains why for moderately overlapping supports the power of the BH procedure decreases and the power of the Bonferroni procedure increases. When peaks overlap by more than 50% the power of both procedures decreases. For  $\gamma = 2$  the behaviour of power is the same. The power is relatively low before overlapping; therefore it increases for moderate overlapping and falls down for large overlapping. The last panel in Figure 6 shows that when peaks are close or have little overlap, small values of  $\gamma$  perform better.

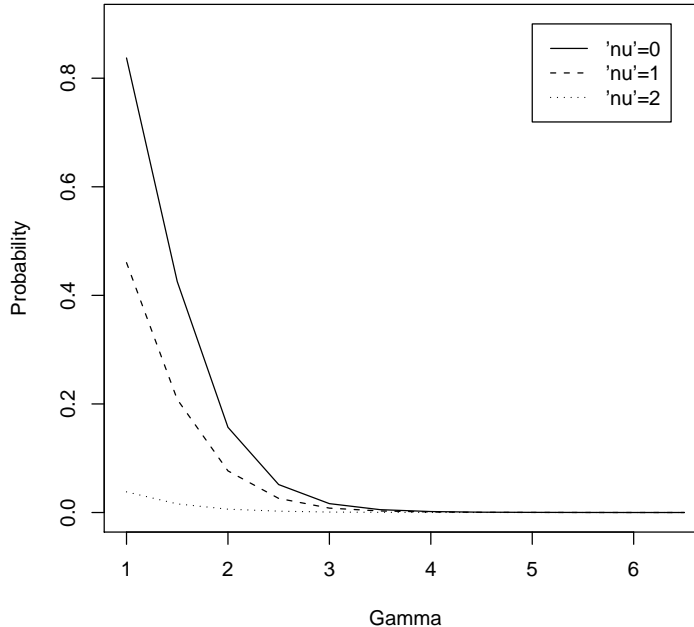


Figure 5: Probability of obtaining more than one local maximum within the same true peak for moderate SNR ( $a = 10$ ).

## 4 Data Example

### 4.1 Data description and analysis

Our data consists of 60 seconds of recordings from an electrode attempting to capture the activity of a single type of neuron in a salamander brain. The recorded signal was digitized at a sampling frequency of 10 KHz, resulting in a sequence of 600,000 measurements equally spaced over time. Data of these kind and over much longer time periods are routinely collected in neuroscience experiments (Baccus and Meister, 2002).

This is a situation where the model of Section 2.1 can be applied directly. Since all peaks of interest correspond to action potentials from the same type of cell, it is reasonable to assume that they all have the same shape. Their intensities, however, vary according to the distances of the cells to the recording electrode. The depolarization (positive) component of the action potential is unimodal. The noise is a mixture of electrical noise and recording of remote cells, exhibiting roughly a large-scale stationary behaviour. Favourably, the SNR is high. For illustration, the (smoothed) data is depicted in Figure 7.

Procedure 2.1 was implemented as follows. For Step 1, we used a Gaussian kernel as in Example 2.5. From the results of Section 3, the choice of kernel width  $\gamma$  is not crucial but it is better if it roughly matches the width of the signal peaks. Since all peaks are assumed to have the same shape, we selected a few ‘obvious’ peaks and estimated their width. Figure 8 shows that a scaled normal density with standard deviation of 1.5 approximates the peaks shape reasonably well.

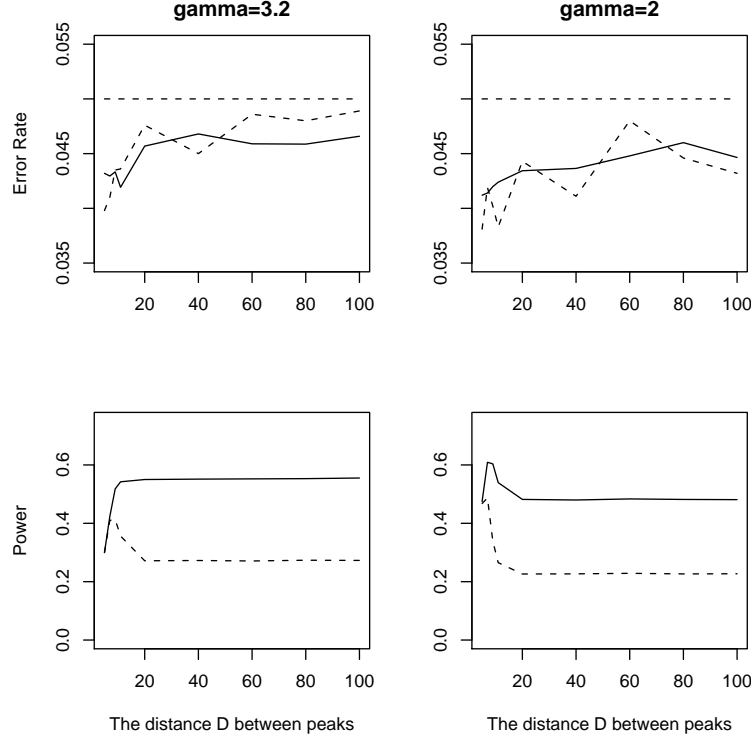


Figure 6: Effect of distance between peaks  $D$  on the BH procedure (solid) and the Bonferroni procedure (dashed).

In Step 2,  $\tilde{m} = 54,452$  local maxima were found. For the heights of these local maxima, the corresponding p-values in Step 3 were computed according to formula (16), plugging in estimates of the spectral moment parameters  $\sigma^2$ ,  $\lambda_2$  and  $\lambda_4$  as described below.

For Step 4, we applied the BH procedure with  $\tilde{m} = 54,452$  and level  $q = 0.01$ , leading to 464 rejections of the null hypothesis. Figure 7 shows the smoothed data and the BH threshold. While many of the peaks could have been found by simple inspection, thanks to the high SNR, the algorithm detected many other weaker peaks. Figure 7(b) shows a zoomed segment of the data, in which two quite different peaks were detected, one strong and one weak.

## 4.2 Estimation of spectral moments

Here we describe how to estimate the noise parameters  $\sigma^2$ ,  $\lambda_2$  and  $\lambda_4$  using a simple method motivated by (9). It is known that the variance of an ergodic process can be consistently estimated by the sample variance. This suggests estimating  $\sigma^2$ ,  $\lambda_2$  and  $\lambda_4$  respectively by the sample variance of the process, the difference process (as an approximation to the derivative), and the difference squared process (as an approximation to the second derivative). However, the smoothed observed process  $x_\gamma(t)$  is expected to contain some signal and not just noise, biasing the sample variance estimators. To reduce the bias, we replace the standard sample variance by a nonparametric estimator that is less sensitive to extreme

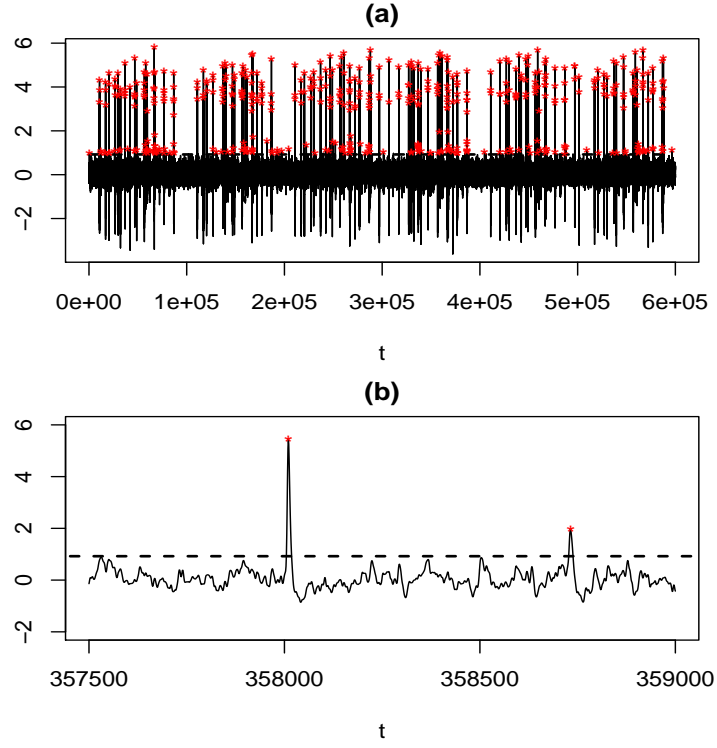


Figure 7: (a) The neural spike data after smoothing with a Gaussian kernel of standard deviation 1.5. (b) Zoom in. In both panels, the stars indicate the detected peaks according to the BH procedure. The dashed line is the BH threshold for rejecting the null hypothesis.

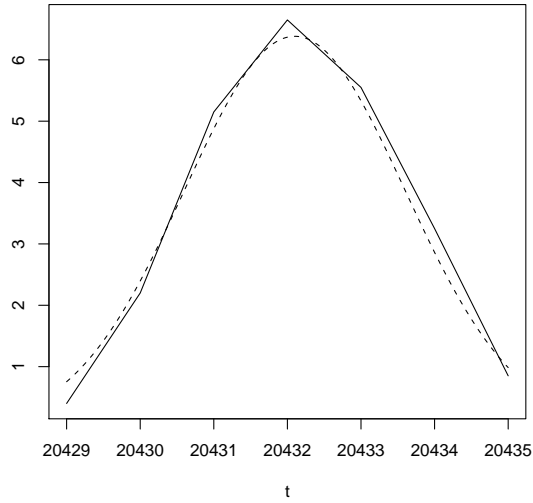


Figure 8: One ‘obvious’ peak (solid) superimposed on a scaled normal density with standard deviation 1.5 (dashed).

values, specifically

$$\begin{aligned}\hat{\sigma}^2 &= \text{MAD}^2 [x_\gamma(t)] = \text{Med}^2 \{|x_\gamma(t_i) - \text{Med} [x_\gamma(t)]|\}, \\ \hat{\lambda}_2 &= \text{MAD}^2 [\Delta x_\gamma(t)], \\ \hat{\lambda}_4 &= \text{MAD}^2 [\Delta^2 x_\gamma(t)],\end{aligned}\tag{31}$$

where  $\Delta x(t)$  for any discretely sampled vector  $x(t) = \{x(t_i)\}_{i=1}^L$  is the sequence of differences  $\{[x(t_{i+1}) - x(t_i)]/(t_{i+1} - t_i)\}_{i=1}^{L-1}$ . These estimators are expected to perform well if the signal is sparse in the sense that it occupies only a small portion of the data.

To evaluate the accuracy of the proposed estimators, we performed a brief simulation in which we compared their performance with the performance of three additional estimation methods. The first method estimates the spectral moments by standard sample variances. The second method is based on the fact that for a stationary zero-mean Gaussian process  $z(t) \in \mathcal{C}^2$  with autocovariance function (ACF)  $c(s) = \text{E}[z(t)z(t+s)]$ , the spectral moments satisfy  $\sigma^2 = c(0)$ ,  $\lambda_2 = -\ddot{c}(0)$  and  $\lambda_4 = c^{(4)}(0)$ . If  $\hat{c}(s)$  is the empirical ACF of the noise process, we fit a polynomial regression  $\hat{c}(s) = \beta_0 + \beta_2 s^2 + \beta_4 s^4 + \varepsilon$  in the neighbourhood of  $s = 0$ , where even powers of  $s$  suffice because the ACF is symmetric around zero. Computing the derivatives of the polynomial at  $s = 0$  leads to the estimators  $\hat{\sigma}^2 = \hat{\beta}_0$ ,  $\hat{\lambda}_2 = -2\hat{\beta}_2$  and  $\hat{\lambda}_4 = 24\hat{\beta}_4$ . The third method estimates  $\sigma^2$  as in (31) and  $\lambda_2$  by the ‘crossing’ estimator suggested by Lindgren (1974),

$$\hat{\lambda}_2 = \hat{\sigma}^2 \frac{2\pi}{T} \frac{1}{3} \left( N_0(T) + \exp \frac{u^2}{2} N_u(T) + \exp \frac{u^2}{2} N_{-u}(T) \right),$$

where  $u = 2\hat{\sigma}/3$  and  $N_u(T)$  is the number of upcrossings of the process  $x_\gamma(t)$  of the level  $u$  in  $[0, T]$ . The estimation of  $\lambda_4$  is done in the same way from the process of differences  $\Delta x_\gamma(t)$ .

In the simulation, two discrete sequences of length 10,000 were generated. The first one contained only white Gaussian noise and the second one contained white Gaussian noise plus 17 equally shaped peaks with heights all equal to 2, corresponding to high SNR. Mimicking the data, both sequences were smoothed with a Gaussian kernel of standard deviation 1.5. The exact moments  $\sigma^2$ ,  $\lambda_2$  and  $\lambda_4$  were computed via (15). The estimated moments by all the methods were averaged over 2000 replications. Table 2 summarizes the results.

Returning to the data, after subtracting the overall mean of 0.0537, the obtained estimates via (31) were  $\hat{\sigma}^2 = 0.0587$ ,  $\hat{\lambda}_2 = 0.0019$  and  $\hat{\lambda}_4 = 0.0006$ . P-values were then computed for the mean-subtracted data via (10) using these estimated moments, leading to the results described earlier in Section 4.1.

## 5 Discussion

In this paper, we have used the heights of local maxima after smoothing as test statistics for identifying unimodal peaks in the presence of Gaussian stationary noise. It was shown that the procedure provides strong control of FWER and FDR asymptotically as both the SNR and length of the sequence tend to infinity, with the length of the sequence allowed to grow exponentially faster than the SNR. Simulations showed that the algorithm is powerful and that a matched filter principle applies where the optimal smoothing bandwidth is close to the width of the peaks to be detected.

Noise only					
	Formula (15)	MAD	Var	ACF	Lindgren
$\sigma^2$	0.188	0.188 (0.007)	0.188 (0.005)	0.160 (0.005)	0.188 (0.007)
$\lambda_2$	0.042	0.040 (0.001)	0.040 (0.001)	0.014 (0.0004)	0.032 (0.001)
$\lambda_4$	0.009	0.023 (0.001)	0.023 (0.001)	0.002 (0.0001)	0.016 (0.001)

Noise plus sparse signal					
	Formula (15)	MAD	Var	ACF	Lindgren
$\sigma^2$	0.188	0.193 (0.007)	0.201 (0.005)	0.171 (0.005)	0.193 (0.007)
$\lambda_2$	0.042	0.040 (0.001)	0.041 (0.001)	0.015 (0.005)	0.032 (0.001)
$\lambda_4$	0.009	0.024 (0.001)	0.024 (0.001)	0.002 (0.0001)	0.016 (0.001)

Table 2: Simulation results: exact and estimated spectral moments by four different methods. Standard deviations are in parentheses.

The most critical assumptions for the theoretical results presented are that the noise process is stationary ergodic Gaussian and that the signal peaks have equal shape and are unimodal with compact support. The Gaussianity assumption was chosen because it enabled using a closed formula for computing the p-values associated with the heights of local maxima. For non-Gaussian noise, p-values could be computed via simulation and we expect that the error control properties should be preserved, although this does not follow directly from the proofs presented here.

The assumption of compact support for the signal peaks is necessary for the concept of true and false detection to be well defined. The unimodality assumption makes local maxima good representatives of true peaks. This is formally true asymptotically for high SNR, as the probability that a true peak is represented by one and only observed local maximum tends to one. Besides its practical interpretation, this property is also helpful technically in the proof of FDR control. We do not find the assumption of asymptotically high SNR restrictive in the sense that the search space is allowed to grow exponentially faster. Another way of seeing this is that the SNR need only grow logarithmically with respect to the search space. It was shown in the simulations that moderate SNR suffices for good performance.

The assumption that the peaks have equal shape is technically convenient as it allows reducing the asymptotic analysis of many peaks to the analysis of any single one of them. It also simplifies the concept of an optimal bandwidth, as it is the same for all peaks. This assumption is also a realistic one in many practical situations, such as the neural spikes example presented.

Notice that there is no assumption of sparsity of peaks in the theoretical results. The simulations showed that the error rates and power do not suffer much even if the peaks have some overlap. Sparsity is not needed as long as the spectral moments of the noise process are known. However, sparsity is needed so that these moments can be estimated from data, as suggested in the data analysis section.

A technical issue to consider in practice is that the theory was developed for continuous processes, while in simulations and real data the observations are obtained in a regular discrete grid. The theoretical results carry through approximately if the grid is fine enough, but break down if the smoothing parameter  $\gamma$  is less than the grid spacing. As it is well known in signal processing, sampling before smoothing, inevitable in practice in Step 1

of the procedure, introduces aliasing. Convolution of a sampled discrete sequence with a sampled discrete kernel is not equivalent to sampling the convolution of a continuous process with a continuous kernel, and the approximation gets worse as the smoothing parameter  $\gamma$  gets close to 1. For this reason, we did not include values of  $\gamma$  between 0 to 1 in our simulation study, and we generally do not recommend to smooth using a value of  $\gamma$  that is too close to the grid spacing, since for that case the theory is not valid.

From a broad perspective, we see the methods presented in this paper not only as a solution to the peak detection problem but as an extension of multiple testing paradigms to temporal and spatial domains. While FWER methods for random fields and have been well established, particularly in neuroimaging (Worsley et al., 2004), similar extensions of FDR methods have proven to be more difficult Pacifico et al. (2004a,b). Other related approaches include testing for a spatial signal in the wavelet transform domain (Shen et al., 2002) and FDR for pre-defined spatial clusters (Benjamini and Heller, 2007). Standard FDR methods can be applied in a discretized spatial domain (Genovese et al., 2002) but are inherently designed for discrete units and ignore the spatial structure of the data. Instead, it has been argued by Chumbley and Friston (2009) that in the case of smooth spatial signals, inference should be about topological features, such as cluster volume or peak height. These authors and others (Zhang et al., 2009) have focused on cluster volume. Our work has focused on peak height.

Potential extensions of this work include exploration of the role of sparsity in the estimation of the noise parameters, as well as adapting the procedure to situations where the observed process is not Gaussian, where the peaks do not have a constant width, or where the domain is two- or three-dimensional, as in medical image analysis.

## 6 Proofs

### 6.1 Gaussian autocorrelation model

**Lemma 6.1.** *Let  $w_\nu(t) = (1/\nu)\phi(t/\nu)$ , where  $\phi(t)$  is the standard normal density.*

1. For  $\gamma, \nu > 0$ ,

$$w_\gamma(t) * w_\nu(t) = w_\xi(t), \quad \text{with } \xi = \sqrt{\gamma^2 + \nu^2}.$$

2. Let  $\dot{w}_\xi(t)$  and  $\ddot{w}_\xi(t)$  denote the first and second derivatives of  $w_\xi(t)$  with respect to  $t$ . Then

$$\int_{-\infty}^{\infty} [w_\xi(t)]^2 dt = \frac{1}{2\sqrt{\pi}\xi}, \quad \int_{-\infty}^{\infty} [\dot{w}_\xi(t)]^2 dt = \frac{1}{4\sqrt{\pi}\xi^3}, \quad \int_{-\infty}^{\infty} [\ddot{w}_\xi(t)]^2 dt = \frac{3}{8\sqrt{\pi}\xi^5}.$$

*Proof.*

1. Let  $X \sim N(0, \gamma^2), Y \sim N(0, \nu^2)$  with respective densities  $w_\gamma(t)$  and  $w_\nu(t)$ . Then  $X + Y \sim N(0, \gamma^2 + \nu^2)$  with density  $w_\gamma(t) * w_\nu(t) = w_\xi(t)$ ,  $\xi = \sqrt{\gamma^2 + \nu^2}$ .
2. Let  $w_\xi^{(k)}(t)$  denote the  $k$ -th derivative of  $w_\xi(t)$  and let  $H_k(t)$  denote the  $k$ -th Hermite polynomial. Then

$$\int_{-\infty}^{\infty} [w_\xi^{(k)}(t)]^2 dt = \int_{-\infty}^{\infty} \left[ \frac{(-1)^k}{\xi^k} H_k\left(\frac{t}{\xi}\right) \frac{1}{\xi} \phi\left(\frac{t}{\xi}\right) \right]^2 dt = \frac{1}{\xi^{2k+2}} \int_{-\infty}^{\infty} H_k^2\left(\frac{t}{\xi}\right) \phi^2\left(\frac{t}{\xi}\right) dt.$$

But

$$\phi^2\left(\frac{t}{\xi}\right) = \frac{1}{\sqrt{2\pi}}\phi\left(\frac{\sqrt{2}t}{\xi}\right).$$

Thus replacing in the integral and making the change of variable  $x = \sqrt{2}t/\xi$ , we obtain

$$\int_{-\infty}^{\infty} [w_{\xi}^{(k)}(t)]^2 dt = \frac{1}{2\sqrt{\pi}\xi^{2k+1}} \int_{-\infty}^{\infty} H_k^2\left(\frac{x}{\sqrt{2}}\right) \phi(x) dx.$$

The results of the lemma are obtained by setting in particular  $k = 0$  with  $H_0(x) = 1$ ,  $k = 1$  with  $H_1(x) = x$ , and  $k = 2$  with  $H_2(x) = x^2 - 1$ .

□

### Derivations for Example 2.5 (Gaussian autocorrelation model)

Under the complete null hypothesis, we can write

$$x_{\gamma}(t) = w_{\gamma}(t) * z(t) = w_{\gamma}(t) * \sigma \int_{-\infty}^{\infty} w_{\nu}(t-s) dB(s) = \sigma \int_{-\infty}^{\infty} w_{\xi}(t-s) dB(s) \quad (32)$$

with  $\xi = \sqrt{\gamma^2 + \nu^2}$ , where we have used Lemma 6.1 part (1). By Lemma 6.1 part (2),

$$\begin{aligned} \sigma_{\gamma}^2 &= E[x_{\gamma}^2(t)] = \sigma^2 \int_{-\infty}^{\infty} w_{\xi}^2(t-s) ds = \frac{\sigma^2}{2\sqrt{\pi}\xi} \\ \lambda_2 &= E[\dot{x}_{\gamma}^2(t)] = \sigma^2 \int_{-\infty}^{\infty} \dot{w}_{\xi}^2(t-s) ds = \frac{\sigma^2}{4\sqrt{\pi}\xi^3} \\ \lambda_4 &= E[\ddot{x}_{\gamma}^2(t)] = \sigma^2 \int_{-\infty}^{\infty} \ddot{w}_{\xi}^2(t-s) ds = \frac{3\sigma^2}{8\sqrt{\pi}\xi^5}. \end{aligned}$$

## 6.2 Proof of Theorem 2.10 (Weak control of FWER)

**Lemma 6.2.** *Let  $u_{\text{Bon}}^*$  and  $\tilde{u}_{\text{Bon}}$  be the thresholds defined in (19) and (20), respectively. Then  $|\tilde{u}_{\text{Bon}} - u_{\text{Bon}}^*| \rightarrow 0$  in probability as  $L \rightarrow \infty$ .*

*Proof.* Recall that  $\tilde{m}_0$  is the number of local maxima belonging to the set  $\mathbb{S}_0$  in the segment  $[-L/2, L/2]$ . Denote by  $\tilde{m}_0[0, 1]$  the number of local maxima belonging to the set  $\mathbb{S}_0$  in the unit. Using the notation of Lemma 2.4, we have that by ergodicity,

$$\left| \frac{\tilde{m}_0}{L} - E[\tilde{m}_0[0, 1]] \right| \rightarrow 0$$

in probability as  $L \rightarrow \infty$ , where  $E[\tilde{m}_0[0, 1]]$  does not depend on  $L$ . Since  $\log(\cdot)$  is continuous, the continuous mapping theorem gives that

$$\left| \log \frac{\tilde{m}_0}{L} - \log E[\tilde{m}_0[0, 1]] \right| \rightarrow 0 \quad \implies \quad \left| \log \frac{\tilde{m}_0}{\alpha} - \log \frac{E[\tilde{m}_0]}{\alpha} \right| \rightarrow 0,$$

where we have used the additive property of the logarithm.

Define now the monotone increasing function  $\psi_{\gamma} : \mathbb{R}^+ \rightarrow \mathbb{R}$  by  $\psi_{\gamma}(x) = F_{\gamma}^{-1}(1 - e^{-x})$ ,  $x > 0$ , where  $\kappa > 0$  is a constant. The function  $\psi_{\gamma}(x)$  is Lipschitz continuous for all  $x > 1$  because its derivative  $d\psi_{\gamma}(x)/dx = e^{-x}/\dot{F}_{\gamma}[\psi_{\gamma}(x)]$  is bounded for all  $x > 1$ . Hence,

$$\left| \psi_{\gamma}\left(\log \frac{\tilde{m}_0}{\alpha}\right) - \psi_{\gamma}\left(\log \frac{E[\tilde{m}_0]}{\alpha}\right) \right| \rightarrow 0,$$



implying that

$$\left| F_\gamma^{-1} \left( 1 - \frac{\alpha}{\tilde{m}_0} \right) - F_\gamma^{-1} \left( 1 - \frac{\alpha}{E[\tilde{m}_0]} \right) \right| \rightarrow 0$$

as  $L \rightarrow \infty$ . Multiplying by  $\sigma_\gamma$  gives the result.  $\square$

### Proof of Theorem 2.10.

1. By Proposition 2.3 and Lemma 2.4,

$$\begin{aligned} \text{FWER}(u_{\text{Bon}}^*) &= \mathbb{P}\{V(u_{\text{Bon}}^*) \geq 1\} \leq \mathbb{E}[V(u_{\text{Bon}}^*)] = LE[\tilde{m}(u_{\text{Bon}}^*; [0, 1])] \\ &= LE[\tilde{m}(-\infty; [0, 1])] \frac{\mathbb{E}[\tilde{m}(u_{\text{Bon}}^*; [0, 1])]}{\mathbb{E}[\tilde{m}(-\infty; [0, 1])]} = \mathbb{E}[\tilde{m}_0] F_\gamma(u_{\text{Bon}}^*) \end{aligned}$$

by the Lemma 2.4, part (2). Setting  $F_\gamma(u_{\text{Bon}}^*) = \alpha/\mathbb{E}[\tilde{m}_0]$  gives that  $\text{FWER}(u_{\text{Bon}}^*) \leq \alpha$ .

2. Write

$$\text{FWER}(\tilde{u}_{\text{Bon}}) = \mathbb{P} \left\{ \tilde{T} \neq \emptyset \text{ and } \max_{t \in \tilde{T}} x_\gamma(t) > u_{\text{Bon}}^* + (\tilde{u}_{\text{Bon}} - u_{\text{Bon}}^*) \right\}$$

and apply the fact that for any two random variables  $X, Y$  and any two constants  $c, \varepsilon$ :

$$P(X > c + \varepsilon) - P(|Y| > \varepsilon) \leq P(X > Y + c) \leq P(X > c - \varepsilon) + P(|Y| > \varepsilon).$$

Taking  $X = \max_{t \in \tilde{T}} x_\gamma(t)$ ,  $Y = \tilde{u}_{\text{Bon}} - u_{\text{Bon}}^*$  and  $c = u_{\text{Bon}}^*$ ,

$$\text{FWER}(\tilde{u}_{\text{Bon}}) \leq \mathbb{P} \left\{ \tilde{T} \neq \emptyset \text{ and } \max_{t \in \tilde{T}} x_\gamma(t) > u_{\text{Bon}}^* - \varepsilon \right\} + \mathbb{P} \left\{ \tilde{T} \neq \emptyset \text{ and } |\tilde{u}_{\text{Bon}} - u_{\text{Bon}}^*| > \varepsilon \right\}$$

The second summand goes to 0 in probability as  $L \rightarrow \infty$  by Lemma 6.2. For the first summand, we follow a similar argument to the proof of part (1):

$$\mathbb{P} \left\{ \tilde{T} \neq \emptyset \text{ and } \max_{t \in \tilde{T}} x_\gamma(t) > u_{\text{Bon}}^* - \varepsilon \right\} \leq \mathbb{E}[\tilde{m}_0] F_\gamma(u_{\text{Bon}}^* - \varepsilon) = \alpha \frac{F_\gamma(u_{\text{Bon}}^* - \varepsilon)}{F_\gamma(u_{\text{Bon}}^*)}$$

but the last fraction goes to 1 as  $L \rightarrow \infty$ .  $\square$

### 6.3 Proof of Theorem 2.11 (Strong control of FWER)

**Lemma 6.3.** *Assume the model of Section 2.1 and the procedure of Section 2.2. Then as  $L \rightarrow \infty$  and  $a_j \rightarrow \infty, \forall j$  such that  $\forall j, La_j \phi(Ka_j) \rightarrow 0$ , for any constant  $K$ ,*

1. *The expected number of local maxima in the transition region  $\mathbb{S}_{1,\gamma} \setminus \mathbb{S}_1 = \mathbb{S}_0 \setminus \mathbb{S}_{0,\gamma}$  tends to 0:*

$$\mathbb{E} \left[ \#\{t \in \tilde{T} \cap (\mathbb{S}_{1,\gamma} \setminus \mathbb{S}_1)\} \right] = \mathbb{E} \left[ \#\{t \in \tilde{T} \cap (\mathbb{S}_0 \setminus \mathbb{S}_{0,\gamma})\} \right] \rightarrow 0.$$

2. *The probability of obtaining a local maximum in the transition region tends to 0:*

$$\mathbb{P} \left( \#\{t \in \tilde{T} \cap (\mathbb{S}_0 \setminus \mathbb{S}_{0,\gamma})\} \geq 1 \right) \rightarrow 0$$

*Proof.*

1. The expected number of local maxima in any set  $T \subset [-L/2, L/2]$  can be computed by the Kac-Rice formula:

$$\begin{aligned}
& \mathbb{E}[\#\{t \in T : \dot{x}_\gamma(t) = 0, \ddot{x}_\gamma(t) < 0\}] \\
&= \int_T p(\dot{x}_\gamma(t) = 0) \int_{-\infty}^0 |y| p(\ddot{x}_\gamma(t) = y) dy dt \\
&= \int_T p(\dot{z}_\gamma(t) = -\dot{\mu}_\gamma(t)) \int_0^\infty y p(\ddot{z}_\gamma(t) = -y - \ddot{\mu}_\gamma(t)) dy dt, \tag{33}
\end{aligned}$$

where  $p(\cdot)$  denotes probability density. Recall that  $\ddot{z}_\gamma(t) \sim N(0, \lambda_{4,\gamma})$ . The inner integral in (33) is

$$\begin{aligned}
& \int_0^\infty y p(\ddot{z}_\gamma(t) = -y - \ddot{\mu}_\gamma(t)) dy = \int_0^\infty y \frac{1}{\sqrt{\lambda_{4,\gamma}}} \phi\left(\frac{-y - \ddot{\mu}_\gamma(t)}{\sqrt{\lambda_{4,\gamma}}}\right) dy \\
&= \int_{\ddot{\mu}_\gamma(t)/\sqrt{\lambda_{4,\gamma}}}^\infty \left[\sqrt{\lambda_{4,\gamma}} z - \ddot{\mu}_\gamma(t)\right] \phi(z) dz \\
&= -\ddot{\mu}_\gamma(t) + \sqrt{\lambda_{4,\gamma}} \phi\left(\frac{\ddot{\mu}_\gamma(t)}{\sqrt{\lambda_{4,\gamma}}}\right) + \ddot{\mu}_\gamma(t) \Phi\left(\frac{\ddot{\mu}_\gamma(t)}{\sqrt{\lambda_{4,\gamma}}}\right). \tag{34}
\end{aligned}$$

The set  $T$  can be decomposed as the union  $T = T^+ \cup T^-$ , such that  $\ddot{\mu}_\gamma(t) > 0, \forall t \in T^+$  and  $\ddot{\mu}_\gamma(t) < 0, \forall t \in T^-$ . Using the inequality

$$\frac{x}{1+x^2} \phi(x) < 1 - \Phi(x) < \frac{\phi(x)}{x}, \quad x > 0, \tag{35}$$

for  $t \in T^+$ , the expression in (34) is bounded by

$$0 \leq \ddot{\mu}_\gamma(t) \left[ \frac{\sqrt{\lambda_{4,\gamma}}}{\ddot{\mu}_\gamma(t)} \phi\left(\frac{\ddot{\mu}_\gamma(t)}{\sqrt{\lambda_{4,\gamma}}}\right) + \Phi\left(\frac{\ddot{\mu}_\gamma(t)}{\sqrt{\lambda_{4,\gamma}}}\right) - 1 \right] \leq \phi\left(\frac{\ddot{\mu}_\gamma(t)}{\sqrt{\lambda_{4,\gamma}}}\right) \frac{\sqrt{\lambda_{4,\gamma}}}{1 + \ddot{\mu}_\gamma^2(t)/\lambda_{4,\gamma}}. \tag{36}$$

For  $t \in T^-$  the sum of the last two terms of the expression in (34) is bounded by

$$\begin{aligned}
0 &\leq \sqrt{\lambda_{4,\gamma}} \phi\left(\frac{\ddot{\mu}_\gamma(t)}{\sqrt{\lambda_{4,\gamma}}}\right) + \ddot{\mu}_\gamma(t) \Phi\left(\frac{\ddot{\mu}_\gamma(t)}{\sqrt{\lambda_{4,\gamma}}}\right) \\
&= -\ddot{\mu}_\gamma(t) \left[ \frac{\sqrt{\lambda_{4,\gamma}}}{-\ddot{\mu}_\gamma(t)} \phi\left(\frac{-\ddot{\mu}_\gamma(t)}{\sqrt{\lambda_{4,\gamma}}}\right) - 1 + \Phi\left(\frac{-\ddot{\mu}_\gamma(t)}{\sqrt{\lambda_{4,\gamma}}}\right) \right] \\
&\leq \phi\left(\frac{-\ddot{\mu}_\gamma(t)}{\sqrt{\lambda_{4,\gamma}}}\right) \frac{\sqrt{\lambda_{4,\gamma}}}{1 + \ddot{\mu}_\gamma^2(t)/\lambda_{4,\gamma}}. \tag{37}
\end{aligned}$$

Taking  $T = \mathbb{S}_{1,\gamma} \setminus \mathbb{S}_1 = \cup_{j=1}^{J_L} S_{j,\gamma} \setminus S_j$  and decomposing each transition region as  $S_{j,\gamma} \setminus S_j = (S_{j,\gamma} \setminus S_j)^+ \cup (S_{j,\gamma} \setminus S_j)^-$  and replacing in (33) we get

$$\begin{aligned}
0 &\leq \mathbb{E}[\#\{t \in \mathbb{S}_{1,\gamma} \setminus \mathbb{S}_1 : \dot{x}_\gamma(t) = 0, \ddot{x}_\gamma(t) < 0\}] \\
&= J_L \left\{ \int_{(S_{j,\gamma} \setminus S_j)^+} p(\dot{z}_\gamma(t) = -\dot{\mu}_\gamma(t)) \left[ -\ddot{\mu}_\gamma(t) + \sqrt{\lambda_{4,\gamma}} \phi\left(\frac{\ddot{\mu}_\gamma(t)}{\sqrt{\lambda_{4,\gamma}}}\right) + \ddot{\mu}_\gamma(t) \Phi\left(\frac{\ddot{\mu}_\gamma(t)}{\sqrt{\lambda_{4,\gamma}}}\right) \right] dt \right. \\
&\quad + \int_{(S_{j,\gamma} \setminus S_j)^-} p(\dot{z}_\gamma(t) = -\dot{\mu}_\gamma(t)) \left[ \sqrt{\lambda_{4,\gamma}} \phi\left(\frac{\ddot{\mu}_\gamma(t)}{\sqrt{\lambda_{4,\gamma}}}\right) + \ddot{\mu}_\gamma(t) \Phi\left(\frac{\ddot{\mu}_\gamma(t)}{\sqrt{\lambda_{4,\gamma}}}\right) \right] dt \\
&\quad \left. + \int_{(S_{j,\gamma} \setminus S_j)^-} p(\dot{z}_\gamma(t) = -\dot{\mu}_\gamma(t)) [-\ddot{\mu}_\gamma(t)] dt \right\}. \tag{38}
\end{aligned}$$

The last integral in (38) equals

$$\begin{aligned}
0 &\leq \int_{(S_{j,\gamma} \setminus S_j)^-} p(\dot{z}_\gamma(t) = -\dot{\mu}_\gamma(t)) [-\ddot{\mu}_\gamma(t)] dt \\
&= \int_{(S_{j,\gamma} \setminus S_j)^-} \frac{1}{\sqrt{2\pi\lambda_{2,\gamma}}} \exp\left(-\frac{\dot{\mu}_\gamma(t)^2}{2\lambda_{2,\gamma}}\right) (-\ddot{\mu}_\gamma(t)) dt. \tag{39}
\end{aligned}$$

Since  $\ddot{\mu}_\gamma(t)$  is piecewise continuous, we can express the set  $(S_{j,\gamma} \setminus S_j)^-$  as the finite union of closed intervals  $[c_i, d_i]$ , some to the left of  $\tau_{j,\gamma}$  and some to the right. Note that  $\dot{\mu}_\gamma(t)$  does not change sign for all  $t \in [c_i, d_i]$  and it is a decreasing function, since  $\ddot{\mu}_\gamma(t) < 0$ . Therefore,  $\dot{\mu}_\gamma(c_i) > \dot{\mu}_\gamma(d_i)$ . The integral in (39) can be expressed as the finite sum of integrals of the form

$$\begin{aligned}
&\int_{c_i}^{d_i} \frac{1}{\sqrt{2\pi\lambda_{2,\gamma}}} \exp\left(-\frac{\dot{\mu}_\gamma(t)^2}{2\lambda_{2,\gamma}}\right) [-\ddot{\mu}_\gamma(t)] dt = \int_{\dot{\mu}_\gamma(d_i)/\sqrt{\lambda_{2,\gamma}}}^{\dot{\mu}_\gamma(c_i)/\sqrt{\lambda_{2,\gamma}}} \phi(z) dz \\
&= \Phi\left(\frac{\dot{\mu}_\gamma(d_i)}{\sqrt{\lambda_{2,\gamma}}}\right) - \Phi\left(\frac{\dot{\mu}_\gamma(c_i)}{\sqrt{\lambda_{2,\gamma}}}\right) \leq \phi\left(\frac{\dot{\mu}_\gamma(g_i)}{\sqrt{\lambda_{2,\gamma}}}\right) \frac{a_j[\dot{h}_\gamma(c_i) - \dot{h}_\gamma(d_i)]}{\sqrt{\lambda_{2,\gamma}}} \tag{40}
\end{aligned}$$

where  $g_i = d_i$  if the interval is to the left of  $\tau_{j,\gamma}$  and  $g_i = c_i$  if the interval is to the right of  $\tau_{j,\gamma}$ . Using the above result, the integral in (39) is bounded by:

$$\int_{(S_{j,\gamma} \setminus S_j)^-} p(\dot{z}_\gamma(t) = -\dot{\mu}_\gamma(t)) [-\ddot{\mu}_\gamma(t)] dt \leq \phi\left(\frac{\dot{\mu}_\gamma(g^*)}{\sqrt{\lambda_{2,\gamma}}}\right) \frac{\Delta \dot{h}_\gamma}{\sqrt{\lambda_{2,\gamma}}} a_j \tag{41}$$

where  $g^* = \arg \max \phi(\dot{\mu}_\gamma(g_i)/\sqrt{\lambda_{2,\gamma}})$  and  $\Delta \dot{h}_\gamma = \dot{h}_\gamma(c_i) - \dot{h}_\gamma(d_i)$  is the maximal difference. Note that the bounds in (36) and (37) are the same. Plugging the bounds (36), (37) and (41) in equation (38) we get

$$\begin{aligned}
0 &\leq \mathbb{E}[\#\{t \in \mathbb{S}_{1,\gamma} \setminus \mathbb{S}_1 : \dot{x}_\gamma(t) = 0, \ddot{x}_\gamma(t) < 0\}] \\
&\leq J_L \left\{ \int_{(S_{j,\gamma} \setminus S_j)} p(\dot{z}_\gamma(t) = -\dot{\mu}_\gamma(t)) \left[ \phi\left(\frac{\ddot{\mu}_\gamma(t)}{\sqrt{\lambda_{4,\gamma}}}\right) \frac{\sqrt{\lambda_{4,\gamma}}}{1 + \ddot{\mu}_\gamma^2(t)/\lambda_{4,\gamma}} \right] dt \right. \\
&\quad \left. + \phi\left(\frac{\dot{\mu}_\gamma(g^*)}{\sqrt{\lambda_{2,\gamma}}}\right) \frac{\Delta \dot{h}_\gamma}{\sqrt{\lambda_{2,\gamma}}} a_j \right\}. \tag{42}
\end{aligned}$$

For  $t \in S_{j,\gamma} \setminus S_j, \forall j$ ,  $\mu_\gamma(t) = a_j h_\gamma(t - \tau_{j,\gamma})$ .  $\dot{h}_\gamma(t)$  is bounded away from 0, in particular  $|\dot{h}_\gamma(t)|$  is bounded in  $S_{j,\gamma} \setminus S_j$ . Let  $t^*$  be the point in  $S_{j,\gamma} \setminus S_j$  where  $|\dot{h}_\gamma(t)|$  is minimal.

Continuing (40), the expected number of local maxima in  $\mathbb{S}_{1,\gamma} \setminus \mathbb{S}_1$  is bounded above by

$$\begin{aligned}
& J_L \left[ \frac{1}{\sqrt{\lambda_{2,\gamma}}} \phi \left( \frac{a_j \dot{h}_\gamma(t^*)}{\sqrt{\lambda_{2,\gamma}}} \right) \int_{S_{j,\gamma} \setminus S_j} \phi \left( \frac{\ddot{\mu}_\gamma(t)}{\sqrt{\lambda_4}} \right) \frac{\sqrt{\lambda_{4,\gamma}}}{1 + \ddot{\mu}_\gamma^2(t)/\lambda_4} dt + \phi \left( \frac{a_j \dot{h}_\gamma(g^*)}{\sqrt{\lambda_{2,\gamma}}} \right) \frac{\Delta \dot{h}_\gamma}{\sqrt{\lambda_{2,\gamma}}} a_j \right] \\
& \leq J_L \left[ \sqrt{\frac{\lambda_{2,\gamma}}{\lambda_{2,\gamma}}} \phi \left( \frac{a_j \dot{h}_\gamma(t^*)}{\sqrt{\lambda_{2,\gamma}}} \right) |S_{j,\gamma} \setminus S_j| + \phi \left( \frac{a_j \dot{h}_\gamma(g^*)}{\sqrt{\lambda_{2,\gamma}}} \right) \frac{\Delta \dot{h}_\gamma}{\sqrt{\lambda_{2,\gamma}}} a_j \right] \\
& \leq L \sqrt{\frac{\lambda_{2,\gamma}}{\lambda_{2,\gamma}}} \phi \left( \frac{a_j \dot{h}_\gamma(t^*)}{\sqrt{\lambda_{2,\gamma}}} \right) + \frac{J_L}{L} L a_j \phi \left( \frac{a_j \dot{h}_\gamma(g^*)}{\sqrt{\lambda_{2,\gamma}}} \right) \frac{\Delta \dot{h}_\gamma}{\sqrt{\lambda_{2,\gamma}}}.
\end{aligned}$$

This bound goes to 0 by the lemma's conditions.

2. Immediate from the previous part. □

### Proof of Theorem 2.11.

1. By Proposition 2.3 and Lemma 2.4,

$$\begin{aligned}
& \text{FWER}(u_{\text{Bon}}^*) = \mathbb{P}(V(u_{\text{Bon}}^*) \geq 1) = 1 - \mathbb{P}(V(u_{\text{Bon}}^*) = 0) \\
& = 1 - \mathbb{P} \left( \left[ \#\{t \in \tilde{T} \cap \mathbb{S}_{0,\gamma} : \max x_\gamma(t) > u_{\text{Bon}}^*\} + \#\{t \in \tilde{T} \cap (\mathbb{S}_0 \setminus \mathbb{S}_{0,\gamma}) : \max x_\gamma(t) > u_{\text{Bon}}^*\} \right] = 0 \right) \\
& = 1 - \mathbb{P} \left( V_\gamma(u_{\text{Bon}}^*) = 0 \text{ and } \#\{t \in \tilde{T} \cap (\mathbb{S}_0 \setminus \mathbb{S}_{0,\gamma}) : \max x_\gamma(t) > u_{\text{Bon}}^*\} = 0 \right) \\
& \leq \mathbb{P}(V_\gamma(u_{\text{Bon}}^*) \geq 1) + \mathbb{P} \left( \#\{t \in \tilde{T} \cap (\mathbb{S}_0 \setminus \mathbb{S}_{0,\gamma})\} \geq 1 \right). \tag{43}
\end{aligned}$$

The last inequality holds since  $1 - \mathbb{P}(A \cap B) \leq \mathbb{P}(A^C) + \mathbb{P}(B^C)$  and

$$\mathbb{P} \left( \#\{t \in \tilde{T} \cap (\mathbb{S}_0 \setminus \mathbb{S}_{0,\gamma}) : \max x_\gamma(t) > u_{\text{Bon}}^*\} \geq 1 \right) \leq \mathbb{P} \left( \#\{t \in \tilde{T} \cap (\mathbb{S}_0 \setminus \mathbb{S}_{0,\gamma})\} \geq 1 \right).$$

The second probability in (43) goes to 0 by Lemma 6.3, part (2).

Although the process  $x_\gamma(t)$  is not stationary under the true model, it has the same properties as stationary process on the set  $\mathbb{S}_{0,\gamma}$ , since  $x_\gamma(t) = z_\gamma(t)$  for all  $t \in \mathbb{S}_{0,\gamma}$ . Thus, following the same arguments as in the first part of Theorem 2.3 we get that the first probability in (43) is bounded by  $\alpha$ . The second probability goes to 0 by Lemma 6.3.

2. Let  $\tilde{m}_{0,\gamma}$  be the number of local maxima belonging to the set  $\mathbb{S}_{0,\gamma}$  and let  $\tilde{v}_\gamma = F_\gamma^{-1}(\alpha/\tilde{m}_{0,\gamma})$ . It is clear that  $\tilde{v}_\gamma \leq \tilde{u}_{\text{Bon}}$  since  $\tilde{m}_{0,\gamma} \leq \tilde{m}$ , therefore

$$\begin{aligned}
\text{FWER}(\tilde{u}_{\text{Bon}}) &= \mathbb{P}(V(\tilde{u}_{\text{Bon}}) \geq 1) \leq \mathbb{P}(V(\tilde{v}_\gamma) \geq 1) \\
&\leq \mathbb{P}(V_\gamma(\tilde{v}_\gamma) \geq 1) + \mathbb{P} \left( \#\{t \in \tilde{T} \cap (\mathbb{S}_0 \setminus \mathbb{S}_{0,\gamma})\} \geq 1 \right) \tag{44}
\end{aligned}$$

by an argument similar to that in (43). Following the same arguments as in the proof of Theorem 2.10 part (2), the first probability in (44) is bounded by  $\alpha$  as  $L \rightarrow \infty$ . The second probability goes to 0 by Lemma 6.3. □

## 6.4 Proof of Theorem 2.12 (Control of FDR)

**Lemma 6.4.** Recall from Section 2.1 and 2.2 that  $\mu_\gamma(t) = a_j h_\gamma(t - \tau_{j,\gamma})$  with finite support  $S_{j,\gamma}$  and  $h_\gamma(t)$  has a unique local maximum at  $t = \tau_{j,\gamma}$  that is an interior point of  $S_j \subset S_{j,\gamma}$  and has no other critical points. Let  $0 < \varepsilon < 1$  and  $I_{j,\varepsilon} = [\tau_{j,\gamma} - \varepsilon, \tau_{j,\gamma} + \varepsilon] \subset S_j \subset S_{j,\gamma}$ . Then, as  $a_j \rightarrow \infty$  and  $\varepsilon \rightarrow 0$  such that  $a_j \dot{h}_\gamma(-\varepsilon) \rightarrow \infty$  and  $a_j \dot{h}_\gamma(\varepsilon) \rightarrow -\infty$ ,

1.  $\mathbb{P}(\#\{t \in \tilde{T} \cap I_{j,\varepsilon}\} \geq 1) \rightarrow 1$
2.  $\mathbb{P}(\#\{t \in \tilde{T} \cap I_{j,\varepsilon} : x_\gamma(t) > u\} \geq 1) \rightarrow 1$  for every fixed threshold  $u$ .

*Proof.*

1. The probability that  $x_\gamma(t)$  has some local maxima in  $I_\varepsilon$  is

$$\begin{aligned}
1 &\geq \mathbb{P}(\#\{t \in \tilde{T} \cap I_{j,\varepsilon}\} \geq 1) \geq \mathbb{P}\{\dot{x}_\gamma(\tau_{j,\gamma} - \varepsilon) > 0 \text{ and } \dot{x}_\gamma(\tau_{j,\gamma} + \varepsilon) < 0\} \\
&= \mathbb{P}\{\dot{z}_\gamma(\tau_{j,\gamma} - \varepsilon) > -\dot{\mu}_\gamma(\tau_{j,\gamma} - \varepsilon) \text{ and } \dot{z}_\gamma(\tau_{j,\gamma} + \varepsilon) < -\dot{\mu}_\gamma(\tau_{j,\gamma} + \varepsilon)\} \\
&\geq 1 - [\mathbb{P}(\dot{z}_\gamma(\tau_{j,\gamma} - \varepsilon) \leq -\dot{\mu}_\gamma(\tau_{j,\gamma} - \varepsilon)) + \mathbb{P}(\dot{z}_\gamma(\tau_{j,\gamma} + \varepsilon) \geq -\dot{\mu}_\gamma(\tau_{j,\gamma} + \varepsilon))] \\
&\geq 1 - \left[ \mathbb{P}\left(\inf_{[\tau_{j,\gamma} - \varepsilon, \tau_{j,\gamma}]} \dot{z}_\gamma(t) \leq -\dot{\mu}_\gamma(\tau_{j,\gamma} - \varepsilon)\right) + \mathbb{P}\left(\sup_{[\tau_{j,\gamma}, \tau_{j,\gamma} + \varepsilon]} \dot{z}_\gamma(t) \geq -\dot{\mu}_\gamma(\tau_{j,\gamma} + \varepsilon)\right) \right] \\
&= 1 - \left[ \mathbb{P}\left(\sup_{[\tau_{j,\gamma} - \varepsilon, \tau_{j,\gamma}]} -\dot{z}_\gamma(t) \geq a_j \dot{h}_\gamma(-\varepsilon)\right) + \mathbb{P}\left(\sup_{[\tau_{j,\gamma}, \tau_{j,\gamma} + \varepsilon]} \dot{z}_\gamma(t) \geq -a_j \dot{h}_\gamma(\varepsilon)\right) \right]. \quad (45)
\end{aligned}$$

The probability that the supremum of any differentiable random process,  $f(t)$ , is above  $u$  is bounded by (Adler and Taylor, 2007).

$$\mathbb{P}\left(\sup_{t \in [0, T]} f(t) \geq u\right) \leq \mathbb{P}(f(0) \geq u) + \mathbb{E}[N_u] \triangleq G(u, |T|, \sigma^2),$$

where  $N_u$  is the number of up-crossings by  $f$  of the level  $u$  in the interval  $[0, T]$  and  $\sigma^2$  is the variance of the process. For the stationary Gaussian process  $\dot{z}_\gamma(t)$  with mean 0 and variance  $\lambda_{2,\gamma}$ , applying Kac-Rice formula gives:

$$\mathbb{E}[N_u] = |T| p(\dot{z}_\gamma = u) \int_0^\infty x p(\ddot{z}_\gamma = x) dx = \frac{|T|}{2\pi} \exp\left(-\frac{u^2}{2\lambda_{2,\gamma}}\right) \sqrt{\frac{\lambda_{4,\gamma}}{\lambda_{2,\gamma}}}.$$

In particular,

$$\begin{aligned}
\mathbb{P}\left(\sup_{[\tau_{j,\gamma} - \varepsilon, \tau_{j,\gamma}]} -\dot{z}_\gamma(t) \geq u\right) &\leq 1 - \Phi\left(\frac{u}{\sqrt{\lambda_{2,\gamma}}}\right) + \frac{\varepsilon}{2\pi} \exp\left(-\frac{u^2}{2\lambda_{2,\gamma}}\right) \sqrt{\frac{\lambda_{4,\gamma}}{\lambda_{2,\gamma}}} \\
&\leq \frac{\sqrt{\lambda_{2,\gamma}}}{u} \phi\left(\frac{u}{\sqrt{\lambda_{2,\gamma}}}\right) + \varepsilon \sqrt{\frac{\lambda_{4,\gamma}}{2\pi\lambda_{2,\gamma}}} \phi\left(\frac{u}{\sqrt{\lambda_{2,\gamma}}}\right) \\
&= G(u, \varepsilon, \lambda_{2,\gamma}). \quad (46)
\end{aligned}$$

The bracketed expression in (45) can be bounded by

$$\begin{aligned} 0 &\leq \mathbb{P} \left( \sup_{[\tau_{j,\gamma}-\varepsilon, \tau_{j,\gamma}]} -\dot{z}_\gamma(t) \geq a_j \dot{h}_\gamma(-\varepsilon) \right) + \mathbb{P} \left( \sup_{[\tau_{j,\gamma}, \tau_{j,\gamma}+\varepsilon]} \dot{z}_\gamma(t) \geq -a_j \dot{h}_\gamma(\varepsilon) \right) \\ &\leq G \left( a_j \dot{h}_\gamma(-\varepsilon), \varepsilon, \lambda_{2,\gamma} \right) + G \left( -a_j \dot{h}_\gamma(\varepsilon), \varepsilon, \lambda_{2,\gamma} \right). \end{aligned} \quad (47)$$

Both terms in (47) go to 0. Thus back to (45),  $\mathbb{P}(\#\{t \in \tilde{T} \cap I_{j,\varepsilon}\} \geq 1) \rightarrow 1$ .

2. The probability that at least one local maxima in  $I_{j,\varepsilon}$  exceeds the fixed threshold  $u$  satisfies

$$\begin{aligned} \mathbb{P} \left( \#\{t \in \tilde{T} \cap I_{j,\varepsilon} : x_\gamma(t) > u\} \geq 1 \right) &\geq \mathbb{P} \left( \inf_{I_{j,\varepsilon}} x_\gamma(t) > u \right) = \mathbb{P} \left( \inf_{I_{j,\varepsilon}} [\mu_\gamma(t) + z_\gamma(t)] > u \right) \\ &\geq \mathbb{P} \left( \inf_{I_{j,\varepsilon}} \mu_\gamma(t) + \inf_{I_{j,\varepsilon}} z_\gamma(t) > u \right) = \mathbb{P} \left( \inf_{I_{j,\varepsilon}} z_\gamma(t) > u - a_j \inf_{I_{j,\varepsilon}} h_\gamma(t) \right) \\ &= \mathbb{P} \left( \sup_{I_{j,\varepsilon}} -z_\gamma(t) < a_j \inf_{I_{j,\varepsilon}} h_\gamma(t) - u \right) = 1 - \mathbb{P} \left( \sup_{I_{j,\varepsilon}} -z_\gamma(t) \geq a_j \inf_{I_{j,\varepsilon}} h_\gamma(t) - u \right). \end{aligned}$$

The infimum of  $h_\gamma(t)$  in the range  $I_{j,\varepsilon}$  occurs at one of the edges. Without loss of generality we assume that  $\inf_{I_{j,\varepsilon}} h_\gamma(t) = h_\gamma(\varepsilon)$ . Using the upper bound given in equation (46), applied to the interval  $I_{j,\varepsilon}$  and  $z_\gamma(t)$  rather than  $\dot{z}_\gamma(t)$ , yields

$$\mathbb{P} \left( \#\{t \in \tilde{T} \cap I_{j,\varepsilon} : x_\gamma(t) > u\} \geq 1 \right) \geq 1 - C_\varepsilon(u), \quad (48)$$

where

$$C_\varepsilon(u) = \frac{\sigma_\gamma}{a_j h_\gamma(\varepsilon) - u} \phi \left( \frac{a_j h_\gamma(\varepsilon) - u}{\sigma_\gamma} \right) + 2\varepsilon \sqrt{\frac{\lambda_{2,\gamma}}{2\pi\sigma_\gamma^2}} \phi \left( \frac{a_j h_\gamma(\varepsilon) - u}{\sigma_\gamma} \right).$$

For any constant  $u$  the convergence rate of  $a_j h_\gamma(\varepsilon) - u$  is the same as that of  $a_j h_\gamma(\varepsilon)$ . Therefore,  $C_\varepsilon(u) \rightarrow 0$  as  $a_j \rightarrow \infty$  for any fixed threshold  $u$ .

□

**Lemma 6.5.** *Assume the model of Section 2.1 and the procedure of Section 2.2. Let  $\tilde{m}_{1,\gamma} = \#\{\tilde{T} \cap \mathbb{S}_{1,\gamma}\}$  be the number of local maxima in the set  $\mathbb{S}_{1,\gamma}$  and recall that  $W_\gamma(u) = \#\{t \in \tilde{T} \cap \mathbb{S}_{1,\gamma} : x_\gamma(t) > u\}$  is the number of local maxima in  $\mathbb{S}_{1,\gamma}$  above threshold  $u$ . Under the assumptions of Theorem 2.12,*

1. *The probability to get exactly  $J_L$  local maxima in the set  $\mathbb{S}_{1,\gamma}$ ,  $\mathbb{P}(\tilde{m}_{1,\gamma} = J_L) = \mathbb{P}(\#\{\tilde{T} \cap \mathbb{S}_{1,\gamma}\} = J_L)$  tends to 1.*
2. *The probability to get exactly  $J_L$  local maxima in the set  $\mathbb{S}_{1,\gamma}$  that exceed the threshold  $u$ ,*

$$\mathbb{P}(W_\gamma(u) = J_L) = \mathbb{P} \left( \#\{t \in \tilde{T} \cap \mathbb{S}_{1,\gamma} : x_\gamma(t) > u\} = J_L \right) \rightarrow 1$$

*tends to 1, for any fixed threshold  $u$ .*

3.  *$\tilde{m}_{1,\gamma}/L \rightarrow A_1$  in probability.*

4.  $W_\gamma(u)/\tilde{m}_{1,\gamma} \rightarrow 1$  in probability.

*Proof.*

1. The probability to get exactly  $J_L$  local maxima in the set  $\mathbb{S}_{1,\gamma}$  is

$$\begin{aligned}
1 &\geq \mathbb{P}\left(\#\{t \in \tilde{T} \cap \mathbb{S}_{1,\gamma}\} = J_L\right) = \mathbb{P}\left(\#\{t \in \tilde{T} \cap (\cup_{j=1}^{J_L} S_{j,\gamma})\} = J_L\right) \\
&\geq \mathbb{P}\left(\cap_{j=1}^{J_L} \{\#\{t \in \tilde{T} \cap S_{j,\gamma}\} = 1\}\right) \\
&\geq \mathbb{P}\left(\cap_{j=1}^{J_L} \#\{t \in \tilde{T} \cap I_{j,\varepsilon}\} = 1\right), \quad I_{j,\varepsilon} \subset S_{j,\gamma} \\
&\geq 1 - \sum_{j=1}^{J_L} \left[1 - \mathbb{P}\left(\#\{t \in \tilde{T} \cap I_{j,\varepsilon}\} = 1\right)\right].
\end{aligned} \tag{49}$$

Combining equations (45) and (47), the above expression is bounded below by

$$\begin{aligned}
&1 - \sum_{j=1}^{J_L} \left(G\left(a_j \dot{h}_\gamma(-\varepsilon), \varepsilon, \lambda_{2,\gamma}\right) + G\left(-a_j \dot{h}_\gamma(\varepsilon), \varepsilon, \lambda_{2,\gamma}\right)\right) \\
&\geq 1 - J_L \left(G\left(a^* \dot{h}_\gamma(-\varepsilon), \varepsilon, \lambda_{2,\gamma}\right) + G\left(-a^* \dot{h}_\gamma(\varepsilon), \varepsilon, \lambda_{2,\gamma}\right)\right) \\
&\geq 1 - \frac{J_L}{L} L \left[\frac{\sqrt{\lambda_{2,\gamma}}}{a^* \dot{h}_\gamma(-\varepsilon)} \phi\left(\frac{a^* \dot{h}_\gamma(-\varepsilon)}{\sqrt{\lambda_{2,\gamma}}}\right) + \frac{\sqrt{\lambda_{2,\gamma}}}{-a^* \dot{h}_\gamma(\varepsilon)} \phi\left(\frac{a^* \dot{h}_\gamma(\varepsilon)}{\sqrt{\lambda_{2,\gamma}}}\right)\right] \\
&\quad - L \sqrt{\frac{\lambda_{4,\gamma}}{2\pi\lambda_{2,\gamma}}} \left[\phi\left(\frac{a^* \dot{h}_\gamma(\varepsilon)}{\sqrt{\lambda_{2,\gamma}}}\right) + \phi\left(\frac{a^* \dot{h}_\gamma(-\varepsilon)}{\sqrt{\lambda_{2,\gamma}}}\right)\right],
\end{aligned} \tag{50}$$

where  $a^* = \arg \max\{G(a_j \dot{h}_\gamma(-\varepsilon), \varepsilon, \lambda_{2,\gamma}) + G(-a_j \dot{h}_\gamma(\varepsilon), \varepsilon, \lambda_{2,\gamma})\}$ .

The fastest rate of convergence toward zero, for which the requirement is still valid, is achieved when  $\dot{h}_\gamma(\pm\varepsilon)$  converges to zero as  $a^{-(1-\delta)}$ ,  $0 < \delta < 1$ . The lemma's conditions guarantee that expressions in (50) go to 0 in the least favourable case.

2. Following the same computations as in equation (49) and using the bound (48), the probability to have  $J_L$  truly rejected hypotheses is bounded by

$$\begin{aligned}
1 &\geq \mathbb{P}\left(\#\{t \in \tilde{T} \cap \mathbb{S}_{1,\gamma} : x_\gamma(t) > u\} = J_L\right) \\
&\geq 1 - \left(\sqrt{\sigma_\gamma} \frac{J_L}{a_j h_\gamma(\varepsilon) - u} \phi\left(\frac{a_j h_\gamma(\varepsilon) - u}{\sigma_\gamma}\right) + L \sqrt{\frac{\lambda_{2,\gamma}}{2\pi\sigma_\gamma^2}} \phi\left(\frac{a_j h_\gamma(\varepsilon) - u}{\sigma_\gamma}\right)\right).
\end{aligned}$$

Following the same arguments as those in the last paragraph in the first part of this lemma, the expression in parentheses goes to 0 for any fixed threshold  $u$ .

3. Since

$$\frac{\tilde{m}_{1,\gamma}}{L} = \frac{\tilde{m}_{1,\gamma}}{J_L} \frac{J_L}{L},$$

we need to show that  $\tilde{m}_1/J_L \rightarrow 1$  in probability. For any fixed  $\varepsilon > 0$

$$0 \leq \mathbb{P}\left(\left|\frac{\tilde{m}_{1,\gamma}}{J_L} - 1\right| \geq \varepsilon\right) = \mathbb{P}(|\tilde{m}_{1,\gamma} - J_L| \geq J_L \varepsilon) \leq \mathbb{P}(\tilde{m}_{1,\gamma} \neq J_L) = 1 - \mathbb{P}(\tilde{m}_{1,\gamma} = J_L)$$

since  $\tilde{m}_{1,\gamma}$  and  $J_L$  are integers. The probability to get exactly  $J_L$  local maxima goes to 1 by the part (1) of this lemma.

4. By part (2) of this lemma  $P(W_\gamma(u) = J_L) \rightarrow 1$  in probability, therefore, using the same arguments as in part (3) of this lemma, we get  $W_\gamma(u)/J_L \rightarrow 1$ . Now,

$$\frac{W_\gamma(u)}{\tilde{m}_{1,\gamma}} = \frac{W_\gamma(u)}{J_L} \frac{J_L}{\tilde{m}_{1,\gamma}}.$$

Both fractions go to 1 by the previous parts of this lemma.

□

**Lemma 6.6.** *For any fixed threshold  $u$ , and any positive integer  $J$*

$$\text{FDR}(u) \leq P(W(u) \leq J-1) + \frac{E[V(u)]}{E[V(u)] + J},$$

where  $W(u) = \#\{t \in \tilde{T} \cap \mathbb{S}_1 : x_\gamma(t) > u\}$ .

*Proof.* For simplicity, in this proof we omit the argument  $u$ , since it is fixed.

$$\begin{aligned} \text{FDR} &= E\left(\frac{V}{V+W}\right) = \sum_{v=0}^{\infty} \sum_{w=0}^{\infty} \left(\frac{v}{v+w}\right) P(V=v, W=w) \\ &= \sum_{v=0}^{\infty} \sum_{w=0}^{J-1} \left(\frac{v}{v+w}\right) P(V=v, W=w) + \sum_{v=0}^{\infty} \sum_{w=J}^{\infty} \left(\frac{v}{v+w}\right) P(V=v, W=w) \\ &\leq \sum_{w=0}^{J-1} \sum_{v=0}^{\infty} P(V=v, W=w) + \sum_{v=0}^{\infty} \sum_{w=J}^{\infty} \left(\frac{v}{v+J}\right) P(V=v, W=w) \\ &= \sum_{w=0}^{J-1} P(W=w) + \sum_{v=0}^{\infty} \left(\frac{v}{v+J}\right) P(V=v, W \geq J) \\ &\leq P(W \leq J-1) + \sum_{v=0}^{\infty} \left(\frac{v}{v+J}\right) P(V=v) \\ &= P(W \leq J-1) + E\left(\frac{V}{V+J}\right) \leq P(W \leq J-1) + \frac{E(V)}{E(V)+J}. \end{aligned}$$

The last inequality holds by Jensen's inequality, since  $V/(V+J)$  is a concave function of  $V$  for  $V \geq 0$  and  $J \geq 1$ . □

**Proof of Theorem 2.12.** Let  $R(\tilde{u}_{\text{BH}}) = \tilde{m} - k + 1$  be the number of rejected hypotheses using the threshold  $\tilde{u}_{\text{BH}}$ . Let

$$\tilde{G}(u) = \frac{\#\{x_\gamma(t) > u, \dot{x}_\gamma(t) = 0, \ddot{x}_\gamma(t) < 0\}}{\#\{\dot{x}_\gamma(t) = 0, \ddot{x}_\gamma(t) < 0\}}$$

be the empirical marginal right cumulative distribution function of  $x_\gamma(t)|t \in \tilde{T}$ . Then

$$\tilde{G}(\tilde{u}_{\text{BH}}) = \frac{\#\{x_\gamma(t) > \tilde{u}_{\text{BH}}, \dot{x}_\gamma(t) = 0, \ddot{x}_\gamma(t) < 0\}}{\#\{\dot{x}_\gamma(t) = 0, \ddot{x}_\gamma(t) < 0\}} = \frac{R(\tilde{u}_{\text{BH}})}{\tilde{m}}.$$

Multiplying both sides of this equation by  $\alpha$  we get

$$\alpha \tilde{G}(\tilde{u}_{\text{BH}}) = \alpha \frac{R(\tilde{u}_{\text{BH}})}{\tilde{m}} = F_\gamma(\tilde{u}_{\text{BH}}).$$



This result implies that the BH threshold  $\tilde{u}_{\text{BH}}$  is the largest  $u$  that solves the equation

$$\alpha \tilde{G}(u) = F_\gamma(u). \quad (51)$$

The strategy is to solve equation (51) in the limit when  $L, a_j \rightarrow \infty$ . We first find the limit of  $\tilde{G}(u)$ . Letting  $W_\gamma(u) = \#\{t \in \tilde{T} \cap \mathbb{S}_{1,\gamma} : x_\gamma(t) > u\}$  and  $V_\gamma(u) = \#\{t \in \tilde{T} \cap \mathbb{S}_{0,\gamma} : x_\gamma(t) > u\}$ ,

$$\begin{aligned} \tilde{G}(u) &= \frac{\#\{x_\gamma(t) > u, \dot{x}_\gamma(t) = 0, \ddot{x}_\gamma(t) < 0\}}{\#\{\dot{x}_\gamma(t) = 0, \ddot{x}_\gamma(t) < 0\}} = \frac{V_\gamma(u) + W_\gamma(u)}{\tilde{m}_0 + \tilde{m}_1} \\ &= \frac{V_\gamma(u)}{\tilde{m}_{0,\gamma}} \frac{\tilde{m}_{0,\gamma}}{\tilde{m}_{0,\gamma} + \tilde{m}_{1,\gamma}} + \frac{W_\gamma(u)}{\tilde{m}_{1,\gamma}} \frac{\tilde{m}_{1,\gamma}}{\tilde{m}_{0,\gamma} + \tilde{m}_{1,\gamma}} \end{aligned} \quad (52)$$

Recall that  $x_\gamma(t)$  is ergodic, thus by the weak law of large numbers and Lemma 6.5 part (3),

$$\frac{\tilde{m}_{0,\gamma}}{\tilde{m}_{0,\gamma} + \tilde{m}_{1,\gamma}} = \frac{\tilde{m}_{0,\gamma}/L}{\tilde{m}_{0,\gamma}/L + \tilde{m}_{1,\gamma}/L} \rightarrow \frac{\mathbb{E}[\tilde{m}_{0,\gamma}; [0, 1]]}{\mathbb{E}[\tilde{m}_{0,\gamma}; [0, 1]] + A_1}, \text{ as } L \rightarrow \infty. \quad (53)$$

Following the arguments in Lemma 2.4, for  $L \rightarrow \infty$ ,

$$\frac{V_\gamma(u)}{\tilde{m}_{0,\gamma}} \rightarrow \frac{\mathbb{E}[V_\gamma(u)]}{\mathbb{E}[\tilde{m}_{0,\gamma}]} = F_\gamma(u). \quad (54)$$

Finally, by Lemma 6.5, part (4)

$$\frac{W_\gamma(u)}{\tilde{m}_{1,\gamma}} \rightarrow 1 \quad (55)$$

as  $L \rightarrow \infty$  and  $a_j \rightarrow \infty$  such that  $L\phi(Ka_j^\delta) \rightarrow 0$ .

Combining equations (53), (54) and (55) with Lemma 6.5 part (4) in (52), we obtain

$$\tilde{G}(u) \rightarrow F_\gamma(u) \frac{\mathbb{E}[\tilde{m}_{0,\gamma}; [0, 1]]}{\mathbb{E}[\tilde{m}_{0,\gamma}; [0, 1]] + A_1} + \frac{A_1}{\mathbb{E}[\tilde{m}_{0,\gamma}; [0, 1]] + A_1}.$$

Now replacing  $\tilde{G}(u)$  by its limit in (51), we obtain

$$\alpha \left( F_\gamma(u) \frac{\mathbb{E}[\tilde{m}_{0,\gamma}; [0, 1]]}{\mathbb{E}[\tilde{m}_{0,\gamma}; [0, 1]] + A_1} + \frac{A_1}{\mathbb{E}[\tilde{m}_{0,\gamma}; [0, 1]] + A_1} \right) = F_\gamma(u),$$

leading to the solution

$$F_\gamma(u) = \frac{\alpha A_1}{A_1 + \mathbb{E}[\tilde{m}_{0,\gamma}; [0, 1]](1 - \alpha)}. \quad (56)$$

Note that  $F_\gamma(u)$  is convex for  $u > 0$ , therefore exists a unique solution to equation (56).

Let  $u_{\text{BH}}^*$  be the solution of equation (56).

The FDR at the threshold  $u_{\text{BH}}^*$  is bounded by Lemma 6.6 by

$$\begin{aligned} \text{FDR}(u_{\text{BH}}^*) &\leq \mathbb{P}(W(u_{\text{BH}}^*) \leq J_L - 1) + \frac{\mathbb{E}[V(u_{\text{BH}}^*)]}{\mathbb{E}[V(u_{\text{BH}}^*)] + J_L} \\ &= \mathbb{P}(W(u_{\text{BH}}^*) \leq J_L - 1) + \frac{\mathbb{E}[V_\gamma(u_{\text{BH}}^*)] + \mathbb{E}[\#\{t \in \tilde{T} \cap (\mathbb{S}_0 \setminus \mathbb{S}_{0,\gamma}) : x_\gamma(t) > u_{\text{BH}}^*\}]}{\mathbb{E}[V_\gamma(u_{\text{BH}}^*)] + \mathbb{E}[\#\{t \in \tilde{T} \cap (\mathbb{S}_0 \setminus \mathbb{S}_{0,\gamma}) : x_\gamma(t) > u_{\text{BH}}^*\}] + J_L}, \end{aligned}$$

where we have split  $V_\gamma(u_{\text{BH}}^*)$  into the region  $\mathbb{S}_{0,\gamma}$  and the transition region  $\mathbb{S}_0 \setminus \mathbb{S}_{0,\gamma}$ .

When both  $L$  and  $a_j$  go to infinity such that  $L\phi(Ka_j) \rightarrow 0$ , Lemma 6.5 gives

$$0 \leq \mathbb{E} \left[ \#\{t \in \tilde{T} \cap (\mathbb{S}_0 \setminus \mathbb{S}_{0,\gamma}) : x_\gamma(t) > u_{\text{BH}}^*\} \right] \leq \mathbb{E} \left[ \#\{t \in \tilde{T} \cap (\mathbb{S}_0 \setminus \mathbb{S}_{0,\gamma})\} \right] \rightarrow 0. \quad (57)$$

By (54), the remaining terms of the fraction can be written as

$$\frac{\mathbb{E}[V_\gamma(u_{\text{BH}}^*)]}{\mathbb{E}[V_\gamma(u_{\text{BH}}^*)] + J_L} = \frac{F_\gamma(u_{\text{BH}}^*)\mathbb{E}[\tilde{m}_{0,\gamma}; [0, 1]]L}{F_\gamma(u_{\text{BH}}^*)\mathbb{E}[\tilde{m}_{0,\gamma}; [0, 1]]L + J_L} = \frac{F_\gamma(u_{\text{BH}}^*)\mathbb{E}[\tilde{m}_{0,\gamma}; [0, 1]]}{F_\gamma(u_{\text{BH}}^*)\mathbb{E}[\tilde{m}_{0,\gamma}; [0, 1]] + J_L/L}.$$

Since  $u_{\text{BH}}^*$  solves (56), for  $L \rightarrow \infty$  such that  $J_L/L \rightarrow A_1$  the above expression tends to

$$\frac{\alpha \mathbb{E}[\tilde{m}_{0,\gamma}; [0, 1]]}{\alpha \mathbb{E}[\tilde{m}_{0,\gamma}; [0, 1]] + A_1 + (1 - \alpha)\mathbb{E}[\tilde{m}_{0,\gamma}; [0, 1]]} = \alpha \frac{\mathbb{E}[\tilde{m}_{0,\gamma}; [0, 1]]}{\mathbb{E}[\tilde{m}_{0,\gamma}; [0, 1]] + A_1} \leq \alpha. \quad (58)$$

Combining equations (57), (58) and Lemma 6.5 part (3), we obtain

$$\limsup \text{FDR}(u_{\text{BH}}^*) \leq \alpha$$

as  $L$  and  $a_j$  go to infinity such that  $L\phi(Ka_j^\delta) \rightarrow 0$ .

Recall that the BH threshold  $\tilde{u}_{\text{BH}}$  solves the equation (51), and  $u_{\text{BH}}^*$  solves the equation (56), where the empirical marginal distribution,  $\tilde{G}(u)$ , is replaced by its limit. Since  $F_\gamma(t)$  is continuous

$$F_\gamma(\tilde{u}_{\text{BH}}) \rightarrow F_\gamma(u_{\text{BH}}^*), \quad (59)$$

leading to

$$\limsup \text{FDR}(\tilde{u}_{\text{BH}}) \leq \alpha.$$

□

## 6.5 Power

### Proof of Lemma 2.14.

1. Let us show first that  $F_\gamma(a_j h_\gamma(0))/F_\gamma(u_{\text{Bon}}^*) \rightarrow 0$ . To show this, we bound  $F_\gamma(x)$  below by

$$F_\gamma(x) \geq \sqrt{\frac{2\pi\lambda_{2,\gamma}^2}{\lambda_{4,\gamma}\sigma_\gamma^2}} \phi\left(\frac{x}{\sigma_\gamma}\right) \frac{1}{2} = C_1 \phi\left(\frac{x}{\sigma_\gamma}\right), \quad C_1 = \sqrt{\frac{\pi\lambda_{2,\gamma}^2}{2\lambda_{4,\gamma}\sigma_\gamma^2}}. \quad (60)$$

An upper bound is given by

$$F_\gamma(x) \leq 1 - \Phi\left(\frac{x}{\sigma_\gamma}\right) + \sqrt{\frac{2\pi\lambda_{2,\gamma}^2}{\lambda_{4,\gamma}\sigma_\gamma^2}} \phi\left(\frac{x}{\sigma_\gamma}\right) \leq \phi\left(\frac{x}{\sigma_\gamma}\right) \left[ \sqrt{\frac{2\pi\lambda_{2,\gamma}^2}{\lambda_{4,\gamma}\sigma_\gamma^2}} + \frac{\sigma_\gamma}{x} \right],$$

where the first inequality uses the fact that  $\lambda_{4,\gamma}/\Delta \geq 1/\sigma_\gamma^2$ , and the second follows from (35). Assuming that  $x$  is large enough, say  $x > \sigma_\gamma$ , the upper bound takes the form

$$F_\gamma(x) \leq C_2 \phi\left(\frac{x}{\sigma_\gamma}\right), \quad C_2 = \sqrt{\frac{2\pi\lambda_{2,\gamma}^2}{\lambda_{4,\gamma}\sigma_\gamma^2}} + 1. \quad (61)$$

Thus

$$\frac{F_\gamma(a_j h_\gamma(0))}{F_\gamma(u_{\text{Bon}}^*)} \leq C_2 \frac{1}{\alpha} \frac{\mathbb{E}[\tilde{m}]}{L} L\phi\left(\frac{a_j h_\gamma(0)}{\sigma_\gamma}\right), \quad (62)$$

which goes to 0 by the conditions of the lemma and by the fact that  $E[\tilde{m}]/L = E[\tilde{m}; [0, 1]]$  does not depend on  $L$ .

Define now  $v = F_\gamma(u_{\text{Bon}}^*)$  and  $w = F_\gamma(a_j h_\gamma(0))$ . Inverting the bounds (60) and (61), we get bounds

$$\begin{aligned} C_1 \phi\left(\frac{u_{\text{Bon}}^*}{\sigma_\gamma}\right) &\leq v \leq C_2 \phi\left(\frac{u_{\text{Bon}}^*}{\sigma_\gamma}\right) \\ \ln \frac{C_1}{\sqrt{2\pi}} - \frac{(u_{\text{Bon}}^*)^2}{2\sigma_\gamma^2} &\leq \ln v \leq \ln \frac{C_2}{\sqrt{2\pi}} - \frac{(u_{\text{Bon}}^*)^2}{2\sigma_\gamma^2} \\ 2\sigma_\gamma^2 \left( \ln \frac{C_1}{\sqrt{2\pi}} - \ln v \right) &\leq (u_{\text{Bon}}^*)^2 \leq 2\sigma_\gamma^2 \left( \ln \frac{C_2}{\sqrt{2\pi}} - \ln v \right) \end{aligned}$$

for  $u_{\text{Bon}}^*$  and similarly for  $a_j h_\gamma(0)$ . Thus,

$$0 \leq \frac{(u_{\text{Bon}}^*)^2}{(a_j h_\gamma(0))^2} \leq \frac{\ln(C_2/\sqrt{2\pi}) - \ln(v)}{\ln(C_1/\sqrt{2\pi}) - \ln(w)}.$$

Applying L'Hôpital rule, the limit of the above fraction when  $v$  and  $w$  go to zero is the same as the limit of  $w/v$ , which is 0 by (62).

2. In contrast to the Bonferroni threshold, the asymptotic FDR threshold depends only on the proportion of false null hypotheses. Because of the assumption of asymptotically fixed proportion of true peaks,  $J_L/L \rightarrow A_1$  with  $0 < A_1 < 1$ , the limit of the FDR threshold (56) is fixed and finite. This leads to the result.

□

### Proof of Theorem 2.13.

Let  $u$  be any fixed threshold and let  $\varepsilon$  and  $I_{j,\varepsilon}$  be defined as in Lemma 6.4. Then, by the definition (26) and the bound (48) in Lemma 6.4 ,

$$1 \geq \text{Power}(u) \geq \mathbb{P}[\#\{t \in I_{j,\varepsilon} : x_\gamma(t) > u\} \geq 1] \geq 1 - C_\varepsilon(u)$$

Combining the results of Lemmas 2.14 and 6.4,  $C_\varepsilon(u) \rightarrow 0$  and the power of Bonferroni and BH procedures when using the deterministic thresholds  $u_{\text{Bon}}^*$  and  $u_{\text{BH}}^*$  converges to 1 in probability.

It was proven earlier that the gap between the random and deterministic thresholds of the Bonferroni and BH procedures goes to 0 (see part (2) of Theorems 2.10, 2.11 and (59)). Therefore, the power when using random thresholds  $\tilde{u}_{\text{Bon}}$  and  $\tilde{u}_{\text{BH}}$  converges to 1 as well.

□

## Acknowledgments

The authors thank Pablo Jadzinsky for providing the neural recordings data, and Igor Wigman, Felix Abramovich, and Yoav Benjamini for helpful discussions. This work was partially supported by NIH grant P01 CA134294-01, the Claudia Adams Barr Program in Cancer Research, and the William F. Milton Fund.

## References

- Robert J. Adler and Jonathan E. Taylor. *Random fields and geometry*. Springer, New York, 2007.
- Robert J. Adler, Jonathan E. Taylor, and Keith J. Worsley. Applications of random fields and geometry: Foundations and case studies. Preprint available at <http://web.ee.technion.ac.il/people/adler/publications.html>, 2010.
- Natalia M. Arzeno, Zhi-De Deng, and Chi-Sang Poon. Analysis of first-derivative based QRS detection algorithms. *IEEE Trans Biomed Eng*, 55(2):478–484, 2008.
- Stephen A. Baccus and Markus Meister. Fast and slow contrast adaptation in retinal circuitry. *Neuron*, 36(5):909–919, 2002.
- Yoav Benjamini and Ruth Heller. False discovery rates for spatial signals. *J Amer Statist Assoc*, 102(480):1272–1281, 2007.
- Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Statist Soc B*, 57(1):289–300, 1995.
- Pierpaolo Brutti, Christopher R. Genovese, Christopher J. Miller, Robert C. Nichol, and Larry Wasserman. Spike hunting in galaxy spectra. Technical report, Libera Università Internazionale degli Studi Sociali Guido Carli di Roma, 2007. URL <http://www.stat.cmu.edu/tr/tr828/tr828.html>.
- Justin R. Chumbley and Karl J. Friston. False discovery rate revisited: FDR and topological inference using Gaussian random fields. *Neuroimage*, 44(1):62–70, 2009.
- Harald Cramér and M. Ross Leadbetter. *Stationary and related stochastic processes*. Wiley, New York, 1967.
- David L. Donoho and Iain M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455, 1994.
- David L. Donoho and Iain M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *J Amer Statist Assoc*, 90(432):1200–1224, 1995.
- Christopher R. Genovese, Nicole A. Lazar, and Thomas E. Nichols. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage*, 15: 870–878, 2002.
- Jaroslav Harezlak, Michael C. Wu, Mike Wang, Armin Schwartzman, David C. Christiani, and Xihong Lin. Biomarker discovery for arsenic exposure using functional data. analysis and feature learning of mass spectrometry proteomic data. *Journal of Proteome Research*, 7(1):217–224, 2008.
- Lei M. Li and Terrence P. Speed. Parametric deconvolution of sparse positive spikes. *Ann Statist*, 28(5):1279–1301, 2000.
- Lei M. Li and Terrence P. Speed. Deconvolution of sparse positive spikes. *J Comp Graph Statist*, 13(4):853–870, 2004.

- Georg Lindgren. Spectral moment estimation by means of level crossings. *Biometrika*, 61: 401–418, 1974.
- Jeffrey S. Morris, Kevin R. Coombes, John Koomen, Keith A. Baggerly, and Ryuji Kobayashi. Feature extraction and quantification for mass spectrometry in biomedical applications using the mean spectrum. *Bioinformatics*, 21(9):1764–1775, 2006.
- Michael S. O’Brien, Anthony N. Sinclair, and Stuart M. Kramer. Recovery of a sparse spike train time series by  $l_1$  norm deconvolution. *IEEE Trans Signal Proc*, 42(12):3353–3365, 1994.
- M. Perone Pacifico, C. Genovese, I. Verdinelli, and L. Wasserman. False discovery control for random fields. *J Amer Statist Assoc*, 99(468):1002–1014, 2004a.
- M. Perone Pacifico, C. Genovese, I. Verdinelli, and L. Wasserman. Scan clustering: A false discovery approach. *J Multivar Anal*, 98(7):1441–1469, 2004b.
- William K. Pratt. *Digital image processing*. Wiley, New York, 1991.
- S. O. Rice. Mathematical theory of random noise. *Bell. System Tech. J*, 25:46–156, 1945.
- Xiaotong Shen, Hsin-Cheng Huang, and Noel Cressie. Nonparametric hypothesis testing for a spatial signal. *J Amer Statist Assoc*, 97(460):1122–1140, 2002.
- Marvin Simon. *Digital communication techniques: signal design and detection*. Prentice Hall, Englewood Cliffs, NJ, 1995.
- Robert Tibshirani, Michael Saunders, Saharon Rosset, Ji Zhu, and Keith Knight. Sparsity and smoothness via the fused lasso. *J R Statist Soc B*, 67(1):91–108, 2005.
- Keith J. Worsley, Jonathan E. Taylor, F. Tomaiuolo, and J. Lerch. Unified univariate and multivariate random field theory. *Neuroimage*, 23:S189–195, 2004.
- Yutaka Yasui, Margaret Pepe, Mary Lou Thompson, Bao-Ling Adam, Jr. George L. Wright, Yinsheng Qu, John D. Potter, Marcy Winget, Mark Thornquist, and Ziding Feng. A data-analytic strategy for protein biomarker discovery: profiling of high-dimensional proteomic data for cancer detection. *Biostatistics*, 4(3):449–463, 2003.
- Hui Zhang, Thomas E. Nichols, and Timothy D. Johnson. Cluster mass inference via random field theory. *Neuroimage*, 44:51–61, 2009.
- Yong Zhang, Tao Liu, Clifford A. Meyer, Jérôme Eeckhoutte, David S. Johnson, Bradley E. Bernstein, Chad Nussbaum, Richard M. Myers, Myles Brown, Wei Li, and X. Shirley Liu. Model-based analysis of ChIP-Seq (MACS). *Genome Biology*, 9(9):R137, 2008.